STUDIES ON THE MECHANISMS OF RNA-DRIVEN DNA REPAIR AND MODIFICATION

A Dissertation
Presented to
The Academic Faculty

by

Ying Shen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
December, 2011

STUDIES ON THE MECHANISMS OF RNA-DRIVEN DNA REPAIR AND MODIFICATION

Approved by:

Dr. Francesca Storici, Advisor
School of Biology
Georgia Institute of Technology

Dr. Yuhong Fan
School of Biology
Georgia Institute of Technology

Dr. Kirill Lobachev
School of Biology
Georgia Institute of Technology

Dr. Anita Corbett
Department of Biochemistry
School of Medicine
Emory University

Dr. Yury Chernoff
School of Biology
Georgia Institute of Technology

Date Approved: [Month dd, yyyy]

To my parents and friends

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Dr. Francesca Storici for her guidance, encouragement and support. It has been an honor to be her first Ph.D. student. She has taught me how good science is done. The passion she has for her research motivates me all the time in the Ph.D. pursuit. I appreciate all of her contributions of time, ideas, and funding to make my Ph.D. study productive. I am also thankful for the excellent example she has provided as a woman biologist and professor.

I would like to thank all my committee members, Dr. Kirill Lobachev, Dr. Yury Chernoff, Dr. Yuhong Fan, and Dr. Anita Corbett for their guidance and feedback on my Ph.D. study. I would also like to thank our collaborator Dr. Bernard Weiss for his insight and advice for our collaboration projects.

I would like to thank all the members of the Storici lab: Rekha Pai, Kuntal Mukherjee, Patrick Ruff, Samantha Stuckey, Kyung Duk Koh, Havva Keskin, as well as former and present undergraduate students and rotation students. Especially, I would like to acknowledge Kyung Duk Koh and Katie Ashley for their immense help for my experiments. I appreciate everyone for their effort to make our group full of friendship, good advice and great fun.

Finally, I would like to thank my parents. Their love, encouragement and understanding have always support me through hard times in my life.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF SYMBOLS AND ABBREVIATIONS	xi
SUMMARY	xiv
CHAPTER	
1 Introduction	1
1.1. Genome instability and DNA repair	1
1.2. rNMPs can be incorporated into chromosomal DNA	2
1.3. The role of RNases H in the degradation of RNA/DNA hybrids	4
1.4. Mismatch repair	6
1.5. Application of oligos in the studies of DNA repair and modification	9
1.6. Research goals	17
2 RNA-driven genetic changes in bacteria and human cells	19
2.1. Abstract	20
2.2. Introduction	21
2.3. Materials and methods	24
2.4. Results and discussion	29
2.5. Conclusions	48
2.6. Acknowledgements	50
3 Generation of RNA/DNA hybrids in genomic DNA by transformation using RNA-containing oligonucleotides	51

3.1. Abstract 52

3.2. Protocol 53

3.3. Materials 60

3.4. Representative results 63

3.5. Discussion 68

3.6. Acknowledgements 70

4 Mispaired rNMPs in DNA are mutagenic and are targets of mismatch repair and RNases H 71

4.1. Abstract 72

4.2. Introduction 73

4.3. Materials and methods 75

4.4. Results 80

4.5. Discussion 97

4.6. Acknowledgements 102

5 Conclusions 103

6 Implications 107

APPENDIX A: Supplementary materials of Chapter 4 112

REFERENCES 131

LIST OF TABLES

Page

Table 2.1: DNA sequences of oligos used in this study and part of the GFP locus disrupted by the I-*Sce*I site 28

Table 2.2: Strand-bias targeting of RNA-containing oligos in *E. coli* 34

Table 2.3: Alkali lability of RNA-containing oligos in *E. coli* 34

Table 4.1: The effect of RNase HI (*rnhA*) and RNase HII (*rnhB*) mutations on transformation by rNMP-containing oligos in *E. coli mutS* cells 84

Table 4.2: Tolerance of an rG/dT mispair in *E. coli* 90

Table 4.3: Strong effect of RNase HII and minor effect of MutS in preventing gene correction by an rC/dA mispair in *E. coli* 92

Table 4.4: Competition between mismatch repair and RNase H type 2 functions in the removal of RNA/DNA mispairs in yeast 96

Table A.1: Bacterial and yeast strains used in this study 114

Table A.2: Oligos used in this study 116

Table A.3: Statistical comparisons (P-values) between gene correction frequencies obtained for different oligos in different genetic backgrounds 117

Table A.4: Alkali lability of RNA-containing oligos used in *E. coli* or yeast transformation assays 125

Table A.5: Reversion frequency of a nonsense mutation or a two-base deletion in the yeast *trp5* gene following transformation by rNMP-containing oligos in MMR and RNase H mutant cells 127

Table A.6: Deletion and nonsense mutations of *rnhA* and *rnhB* genes similarly affect gene correction by oligos in the BW1988 and the BW2037 backgrounds 130

LIST OF FIGURES

Page

Figure 1.1: Scheme of rNMPs present in genomic DNA 3

Figure 1.2: Scheme of the methyl-directed MMR in *E. coli* 8

Figure 1.3: Scheme of the MMR in eukaryotes 9

Figure 1.4: Scheme of oligo-directed DNA modification via the Lambda Red recombination system in *E. coli* 11

Figure 1.5: Repair of a DSB by RNA-containing oligos 15

Figure 1.6: Strand bias targeting of RNA-containing oligos and their complementary DNA-only oligos 16

Figure 2.1: Scheme of the *E. coli* chromosomal region with the *lacZ* gene mutated and then repaired by the *LacZ.R6* oligo or the *LacZ.R6 comp* oligo 33

Figure 2.2: Chromosomal DNA modifications mediated by RNA-containing oligos in human cells 37

Figure 2.3: Alkali lability of RNA-containing oligos in human cells 40

Figure 2.4: Detection of GFP bands following PCR from genomic DNA of human green fluorescent cells targeted by the DNA-only oligo *GFP.D* or the RNA-containing oligo *GFP.R6* 41

Figure 2.5: Strand-bias targeting to broken chromosomal GFP in human cells 44

Figure 2.6: Plasmid DNA modifications mediated by RNA-containing oligos in human cells 47

Figure 3.1: Schematic diagram of the defective *trp5* gene and of the *TRP5* allele corrected by the RNA-containing oligo 63

Figure 3.2: Transformation result with the RNA-containing oligo 64

Figure 3.3: Detection of genetic information transfer from the RNA-containing oligo to yeast chromosomal DNA by restriction digestion of the PCR product amplifying the targeted genomic region 65

Figure 3.4: DNA sequencing results showing gene correction by the RNA-containing oligo 66

Figure 3.5: Alkali treatment prevents gene correction by the RNA-containing oligo 67

Figure 4.1: Diagrams and sequences of the *E. coli lacZ* and *rpsL* loci and the yeast *trp5* locus targeted by the RNA-containing oligos 83

Figure 4.2: RNase HII cleavage specificity 87

Figure 6.1: Scheme of transcript RNA-mediated DSB repair 108

Figure A.1: Denaturing polyacrylamide gel showing fragments resulting from RNase HII cleavage and random RNA degradation in the absence of RNase HII 112

Figure A.2: Strand-bias targeting effect of two complementary oligos, Trp5.72D and TRP5.72Dcom is mild in wt and cannot be observed in *msh2* mutant strains 113

LIST OF SYMBOLS AND ABBREVIATIONS

8-oxoG	8-hydroxyguanine
AGS	Aicardi-Goutieres syndrome
BER	base-excision repair
bp	basepair
BSA	bovine serum albumin
cDNA	complementary DNA
cm^3	cubic centimeter
CMV	cytomegalovirus
DAM	deoxyadenine methylase
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
dNMP	deoxyribonucleoside monophosphate
dNTP	deoxyribonucleotide
DSB	double-strand break
dsDNA	double-stranded DNA
dUMP	deoxyuridine monophosphate
GFP	green fluorescent protein
HEK	human embryonic kidney
hr	hour
IDL	insertion/deletion loop
Ig	immunoglobulin
JRNase	junction ribonuclease

kb	kilobase
L	liter
LINE	long interspersed element
M	molar
MAT	mating type
mg	milligram
min	minute
mL	milliliter
mm^3	cubic millimeter
MMR	mismatch repair system
mtDNA	mitochondrial DNA
NER	nucleotide-excision repair
nmole	nanomole
nt	nucleotide
O^6-MeG	O^6-Methylguanine
OGG	8-oxoguanine DNA glycosylase
oligo	oligonucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEI	polyethylenimine
pmole	picomole
pol	polymerase
RNA	ribonucleic acid
RNase H	ribonuclease H
rNMP	ribonucleoside monophosphate

rpm	round per minute
RT	reverse-transcriptase
rUMP	uridine monophosphate
SC-Trp	synthetic complete medium without tryptophan
sec	second
ssDNA	single-stranded DNA
TBE	Tris/Borate/EDTA
Ty	Transposons of yeast
UV	ultraviolet
WT	wild-type
μL	microliter
μM	micromolar

SUMMARY

Our previous studies have demonstrated that RNA can serve as a template for double-strand break (DSB) repair in the yeast *Saccharomyces cerevisiae* using synthetic RNA-containing oligonucleotides (oligos). Following this initial work, we show that the RNA tract of RNA-containing oligos can be copied into DNA to transfer a genetic change at the chromosomal level also in the bacterium *Escherichia coli* and in human cells. Exploiting the use of oligos containing ribonucleoside monophosphates (rNMPs), we developed a molecular approach to generate RNA/DNA hybrids of chosen sequence and structure at the chromosomal level in both yeast and *E. coli* cells. Such technique allows us to study how rNMPs present in the DNA genome of cells are tolerated by cells, what factors recognize and target rNMPs in DNA and to what extent the embedded rNMPs may alter genome integrity. Here we proved that mispaired rNMPs embedded into genomic DNA, if not removed, serve as templates for DNA synthesis during chromosomal replication and produce a genetic change. We discovered that mispaired rNMPs that are embedded in genomic DNA are not only targeted by ribonucleases H (RNases H) but also by the mismatch repair (MMR) system both in yeast and in *E. coli*. Our data reveal novel substrates for the MMR system, and also uncover an unpredicted competition between RNase H and MMR for the RNA/DNA mispairs.

CHAPTER 1

INTRODUCTION

1.1. Genome instability and DNA repair

Genome integrity is under constant attack from numerous exogenous agents, including chemical agents, UV radiation, and ionizing radiation, and from endogenous sources, such as reactive oxygen species, endogenous alkylating agents, DNA single- and double-strand breaks resulting from collapsed DNA replication forks, replication errors, impairment of DNA repair pathways, and error-prone translesion synthesis [1, 2]. The accumulation of DNA damages can lead to many kinds of mutations, such as substitutions, insertions, deletions, or chromosomal rearrangements. Although genome instability can be harmful to cells and organisms, it drives evolution at the molecular level and generates genetic variation, such as immunoglobulin (Ig) diversification [3]. To maintain the genomic integrity and normal function of cells, several DNA repair pathways interplay and provide robust defense for the cells, including direct reversal of base damage, base excision repair (BER), nucleotide-excision repair (NER), mismatch repair (MMR), and double-strand break (DSB) repair [1]. In addition to the major types of DNA damages, such as base oxidation, base alkylation, base hydrolysis, base mismatch and strand break [1, 2] , the incorporation of ribonucleotides other than deoxyribonucleotides during DNA synthesis could be one form of DNA damage which has been overlooked for a long time.

1.2. rNMPs can be incorporated into chromosomal DNA

Numerous studies provide evidence that ribonucleoside monophosphates (rNMPs) could be the most abundant among all the non-canonical nucleotides in genomic DNA (**Figure 1.1**). DNA polymerases prevent the incorporation of rNMPs during DNA synthesis efficiently [4] but not perfectly. Several *in vitro* studies showed that DNA polymerases involved in both DNA replication and repair can insert rNMPs during DNA synthesis assays ([5] and references therein), [6]. The pool of ribonucleotides is ~ 36- to 190-fold in excess over the pool of deoxyribonucleotides, increasing the chance of rNMPs to be stably incorporated into DNA [5]. In experiments involving primer extension using the same ribonucleotide and deoxyribonucleotide concentrations found in yeast cell extracts, the major yeast replicative DNA polymerases, DNA polymerases α (Pol α), β (Pol β) and ε (Pol ε), were all found to stably incorporate rNMPs into DNA [5]. These data raised the possibility that more than 10,000 rNMPs may be incorporated into the yeast nuclear genome during each round of replication [5]. Interestingly, regions of highly transcribed DNA in yeast cells were found to contain a high concentration of dUMP [7], and, possibly, these regions may also be hot spots for rNMP incorporation. In addition, DNA primases have very low fidelity, resulting in the incorporation of mispaired rNMPs, and could potentially leave one or several rNMPs embedded in chromosomal DNA by including deoxyribonucleotides in addition to ribonucleotides in the primer sequence [8, 9]. Finally, oxidative damage of DNA can convert a deoxyribonucleotide in DNA to an rNMP [10].

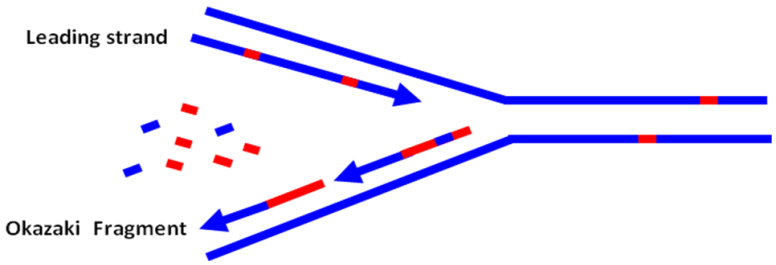


Figure 1.1 Scheme of rNMPs present in genomic DNA. DNA is shown in blue; RNA is in red.

The rNMPs incorporated in the DNA genome could pose a threat to the genome stability for their sensitivity to be cleaved or cause physical distortion of the DNA double helix, due to the presence of the reactive 2'-hydroxyl moiety on the ribose ring [11]. What could the factors in a cell that can recognize and/or remove the rNMPs incorporated in DNA be? Ribonucleases H (RNases H), which have been shown to degrade RNA in an RNA/DNA hybrid *in vitro* [12, 13], can potentially remove rNMPs embedded in DNA to maintain genome stability. In addition, it is possible that DNA repair pathways can also be involved in the process of recognition and removal of rNMPs to maintain genome integrity. The NER repair enzymes recognize bulky distortions in the DNA duplex [14]. It is likely that the bulky ribose with a 2'-hydroxyl group of an rNMP can cause DNA backbone distortion, which serves as a signal for recruiting NER repair enzymes to remove the single-strand DNA segment with the bulky lesion and fill in the gap with the correct dNMPs. The uracil of rUMP can be recognized and removed by DNA glycosylase in BER pathways [15]. Moreover, a base-base mismatch can form when a wrong rNMP is synthesized opposite to a dNMP during DNA replication or a wrong dNMP is synthesized through an rNMP tract. The base-base mismatch from an rNMP/dNMP mispair could potentially be recognized and removed by the DNA MMR pathway. Using

the oligonucleotide (oligo)-driven gene correction assays developed in our lab to form rNMP/dNMP hybrids *in vivo*, we started to investigate if the MMR system can recognize and remove rNMPs in the genome.

1.3. The role of RNases H in the degradation of RNA/DNA hybrids

RNases H are enzymes that hydrolyze the RNA strand of an RNA/DNA heteroduplex [12, 13]. RNases H are involved in DNA replication, transcription, recombination repair, and development[16]. Several *in vitro* biochemical studies indicate that RNases H participate in the removal of RNA primers during lagging-strand synthesis [17, 18]. RNases H are classified into two major families: RNase H type 1 and RNase H type 2 [19]. The type 1 family includes yeast RNase H (1), mammalian RNase H1, prokaryotic RNase HI, and the RNase H domain of reverse transcriptase. The type 2 family includes yeast RNase H (2), mammalian RNase H2, prokaryotic RNases HII and HIII, and archaeal RNase HII [20].

RNases HI/1 and RNases H II/2 have distinct cleavage patterns. RNases HI/1 require a substrate with an RNA stretch containing at least four ribonucleotides in a DNA duplex in order for cleavage to occur [21]. RNases HII exhibit a junction ribonuclease (JRNase) activity, which recognizes an RNA stretch in the RNA-DNA heteroduplex, such as an Okazaki fragment, and cleaves it, leaving a mono-ribonucleotide at the 5' terminus of the RNA-DNA junction. JRNase activity has not been reported for RNase HI/1 and prokaryotic RNases HIII [22]. RNase HII/2 can also recognize and cleave a single ribonucleotide embedded in a DNA duplex, suggesting that RNase H II/2 have a repair

function to cleave a mistakenly incorporated ribonucleotide in a DNA duplex [23]. Moreover, RNase HII from *Chlamydia pneumoniae* can cleave a DNA-rN1-DNA/DNA duplex containing various mismatches around or opposite to the ribonucleotide *in vitro* [24].

RNase H1 and RNase H2 show clear differences in hybrid hydrolysis, indicating that they might have distinct *in vivo* substrates. More detailed analysis of substrate specificity for these two types of enzymes may help to elucidate the so far mostly elusive *in vivo* functions of RNase H1 and RNase H2 [20]. A recent *in vivo* study in yeast strains showed that a mutant form of DNA polymerase ε (*pol2-M644G*) has an increased tendency to incorporate rNMPs into DNA [25]. The gene encoding the catalytic subunit of RNase H2 in yeast is *RNH201*[26]. When the yeast *RNH201*gene was deleted, the *pol2-M644G rnh201* strain appeared to accumulate rNMPs in genomic DNA and had an extended S-phase and an increased rate of spontaneous mutagenesis that included 2–5 bp deletions in repetitive sequences [25]. These data indicate that rNMPs are incorporated during replication *in vivo* and are potentially removed by RNase H2-dependent repair. Failure to remove rNMPs could lead to a replicative stress response and destabilization of the nuclear genome. The deleterious effects could result from DNA distortions, which could affect the binding of proteins to DNA and/or possibly affect the activity of DNA or RNA polymerases in replicating or transcribing through templates containing unrepaired rNMPs.

Eukaryotic RNase H2 is a complex of three different proteins: the catalytic subuit (2A), similar to the monomeric prokaryotic RNase HII, and two other subunits (2B and 2C), which do not have prokaryotic counterparts and clearly known functions, but are necessary for catalysis [20]. RNase H1 has been implicated in mitochondrial DNA (mtDNA) replication during mouse development [27]. Mutations in any of the human RNase H2 subunits can result in Aicardi-Goutieres syndrome (AGS), a severe neurological disorder [28]. The symptons of AGS include progressive microcephaly, spasticity, dystonic posturing, profound psychomotor retardation, death in childhood [28, 29]. So far, very little is known about the *in vivo* substrates of RNases H and which processes RNases H are involved in.

1.4. Mismatch repair

The DNA mismatch repair (MMR) machinery is a highly conserved genome surveillance system in bacteria, yeast and mammals that maintains genome integrity [30-32]. MMR is responsible for correcting practically any kind of DNA/DNA misalignments, including base-base mismatches and insertion/deletion loops (IDL). The base-base mismatches are mostly errors introduced by DNA polymerases that escape proofreading or small DNA lesions generated by intracellular metabolism (eg., oxidative stress) and external chemical agents. The IDLs are usually caused by misalignment of the microsatellite sequences due to strand slippage during DNA synthesis, resulting in different number of microsatellite units between the template and the newly synthesized strand [30, 33]. The absence of MMR leads to mutator phenotypes and a predisposition to cancer in humans [32]. Binding by MMR proteins and repair by MMR have a hierarchical pattern for the eight

mismatch pairs: G-T, A-C, A-A, G-G>T-T, T-C, A-G> C-C [34, 35]. To our knowledge, nobody has examined if a mismatch between a DNA base and an RNA base can be recognized by the cellular MMR system.

So far, mainly two types of MMR have been elucidated. One is employed by *Escherichia coli* and closely related bacteria (described in 1.4.1), and the other is employed by eukaryotes and the majority of bacteria (described in 1.4.2) [31].

1.4.1. Methyl-directed MMR in *E. coli*

In *E. coli*, the recognition of mismatched DNA is mediated by a MutS homodimer, which recruits a homodimer of MutL (**Figure 1.2**). Then the ternary complex of MutS-MutL and DNA activates the endonuclease activity of MutH in an ATP-dependent manner [30]. Since the deoxyadenine methylase (DAM), which methylates adenines in the GATC sites lags behind the replication fork by ~2 minutes, the newly synthesized strand is transiently unmethylated. The MutS-MutL-activated MutH uses this time window to cleave a hemi-methylated GATC site located within about ~1kb of the error, creating a nick for the entry point of the excision reaction [36, 37]. The UvrD helicase is loaded to the nick to unwind the error-containing DNA from the template [38]. The DNA-binding protein SSB binds to the single-strand template to stabilize it, and exonucleases digest the unwound DNA to remove the error either in a 5'$\rightarrow$ 3' direction or 3'$\rightarrow$ 5' direction depending on the location of the nick relative to the mismatch [39]. DNA polymerase III resynthesizes the strand to fill in the gap and the repair is completed when DNA ligase seals the remaining nick.

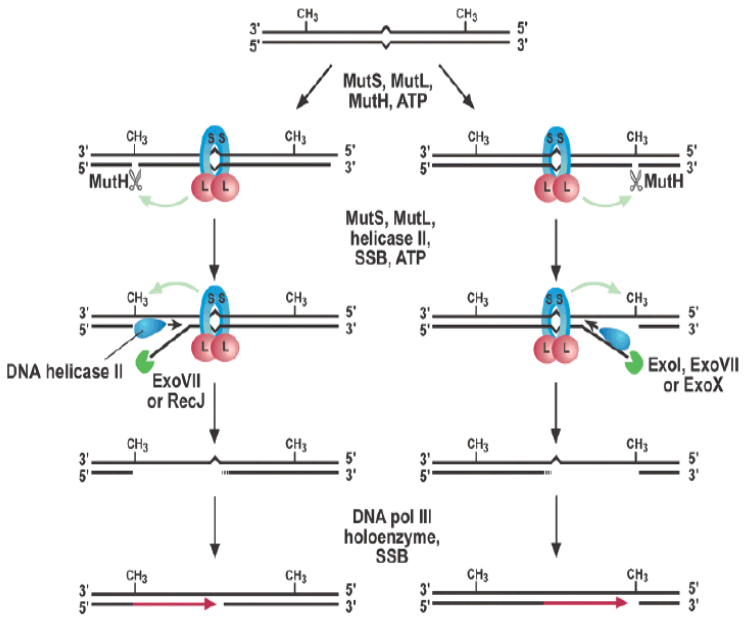


Figure 1.2 Scheme of the methyl-directed MMR in *E. coli* (from [40])

1.4.2. Mismatch recognition in eukaryotes

In eukaryotes, the MMR has features in common with that in *E. coli*, but is more complex (**Figure 1.3**). MutS homologs (Msh), Msh2, Msh3 and Msh6 participate in MMR in the form of heterodimers [30]. Msh2-Msh6 (MutSα) initiates the repair of base-base mismatches and 1-2 base IDLs [41-43], whereas Msh2-Msh3 (MutSβ) initiates the repair of larger IDLs containing up to 16 extra nucleotides in one strand [44]. The heterodimeric MutL homolog, MutLα (MLH-PMS2 in humans, Mlh-Pms1 in yeast) is involved in repairing a wide variety of mismatches. Other MutL homologs, yeast MutLβ (Mlh1-Mlh2) and yeast and human MutLγ (MLH1-MLH3) participate in repairing a subset of IDLs. Although homologs of *E. coli* MutS and MutL exist in almost all organisms, no homolog of MutH has been identified in eukaryotes and most bacteria [30, 41]. In eukaryotes, differently from the methyl-directed mechanism in *E. coli*, strand

discontinuity serves as a signal that directs the targeting of MMR to the discontinuous strand. Discontinuities can exist as at the 3'-ends of the newly synthesized leading stand or at the termini of Okazaki fragments of the newly synthesized lagging strand during DNA replication [31]. MutSα recognizes base-base mismatches and MutLα nicks the 5'- or 3'- side of the mismatched base on the discontinuous strand. The resulting DNA segment is excised by the Exo1 exonuclease in cooperation with the single-stranded DNA binding protein Rpa. The DNA strand is resynthesized by DNA polymerase δ and the nick is ligated by DNA ligase I [30, 31].

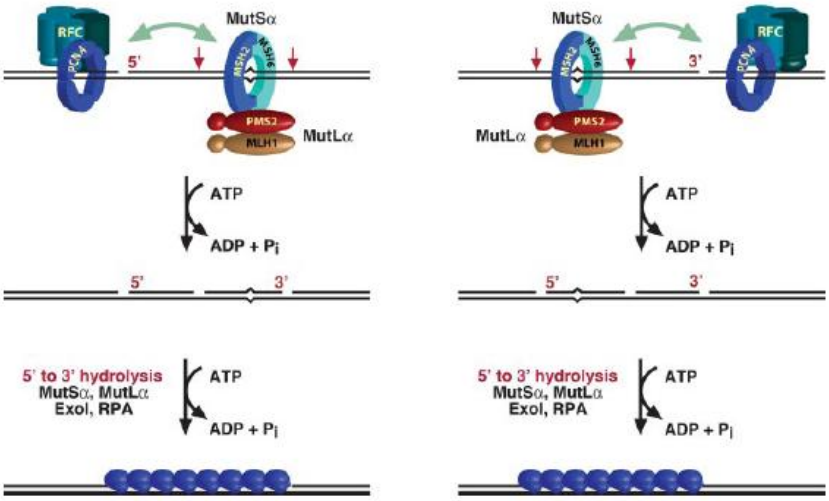


Figure 1.3 Scheme of the MMR in eukaryotes (from [45])

1.5. Application of oligos in the studies of DNA repair and modification

Synthetic oligos are short single-strand DNA or RNA molecules. They are most commonly used as primers for DNA sequencing and amplification, probes for detecting complementary DNA or RNA via molecular hybridization, small interfering RNA, and tools to introduce the site-specific alterations. Particularly, oligos can be synthesized to contain a wide variety of base or sugar modifications to mimic specific nucleic acid

alterations. For example, a site-specifically modified O^6-MeG DNA oligo was used to mimic O6-Methylguanine produced by endogenous and exogenous methylating agents in DNA [46]. It was found that the O^6-MeG largely blocks the elongation of RNA polymerase and generates extensive base misincorporation opposite to O^6-MeG in full length RNA [46]. Using an 8-oxoG containing oligo, it was revealed that 8-oxoGs are primarily repaired by the BER pathway and the repair is initiated by 8-oxoguanine DNA glycosylase (OGG) [47]. Therefore, oligos can serve as important tools for studying the effects of DNA damage and DNA repair mechanisms. In our study we have exploited the use of oligos to study mechanisms of MMR in the repair of RNA/DNA mismatches both in prokaryotic *E. coli* and in eukaryotic *S. cerevisiae* cells.

1.5.1. DNA oligos-directed gene modification

Single-stranded DNA (ssDNA) oligos have been used to edit the chromosomal DNA of different cell types, including *E. coli*, yeast and mammalian cells [48-51], ([52] and references therein). The Red recombination system of bacteriophage Lambda is extensively used in *E.coli* for efficient linear DNA recombination [49]. The Lambda Red system requires all three phage proteins Gam, Exo, and Beta for recombination between double stranded DNA (dsDNA) and the bacterial chromosome [53-55], whereas recombination with ssDNA oligos requires only Beta protein [49]. Beta protein binds to ssDNA, protecting it from degradation by nucleases and promoting the annealing to its complementary ssDNA [56, 57] (**Figure 1.4**). Recombination through the Lambda Red system using ssDNA exhibits high *in vivo* targeting efficiency and is not dependent on RecA in *E. coli*. Beta-dependent ssDNA recombination occurs at the replication fork, in

which the ssDNA oligos corresponding to Okazaki fragments, referred as the lagging-strand oligos, always have higher efficiency for recombination than the leading-strand oligos [49, 55]. Single-strand DNA oligos generating a misalignment with the chromosomal sequence during the Lambda Red recombination process can activate the MMR system to remove the oligo sequence-generating mismatch. Thus, defects in the MMR system can enhance the recombination efficiency of the Lambda Red system by over 100-fold [34].

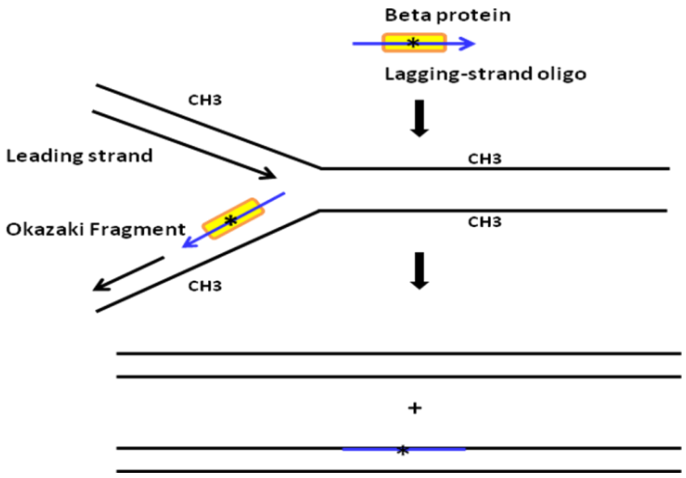


Figure 1.4 Scheme of oligo-directed DNA modification via the Lambda Red recombination system in *E. coli* (modified from [34]).

DNA oligos, homologous to the target sequence except for a single mismatch, were transformed into yeast cells to generate desired mutations on the reporter genes on chromosomal DNA [48, 58]. One limitation of this method is that the identification of genes with the introduced mutations requires a selectable phenotype. To overcome this limitation, an oligo-based genome targeting system, the *delitto perfetto* approach, was developed to provide *in vivo* site-directed DNA modification on almost any chosen chromosomal locus [52, 59]. The *delitto perfetto* includes two steps. The first step is to

integrate a counterselectable reporter (CORE) cassette into a chromosomal locus through the highly effient yeast homologus recombination system. The second step is to transform a designed oligo to replace the integrated CORE cassette. Rad52 is essential in the *delitto perfetto* approach, because both steps require homologous recombination [52]. The mechanism of oligo-mediated DNA modification in yeast cells is not as well characterized as in *E. coli*. The strand bias targeting for two complementary DNA oligos is not obvious in yeast cells [52, 58] (Shen *et al.* in press). In mammalian cells, DNA oligos have also been used to produce single-base alterations to genes on either episomal or chromosomal DNA [50, 51]; however, the oligo-mediated gene targeting is poorly efficient [60].

1.5.2. DSB stimulates DNA modification by DNA oligos

The generation of a DSB increases recombination frequency up to several thousand-fold from bacteria to human cells [61-64]. The break-mediated *delitto perfetto* approach developed in the yeast system is a simple and efficient approach for *in vivo* DNA modification using single-strand DNA oligos. The first step of this approach is to generate a specific DSB at the desired chromosomal location by inserting an inducible site-specific endonuclease (I-*Sce*I) together with its cutting site. The second step is to target the chromosomal locus with designed oligos in such a way that the genomic sequence is modified with no remaining traces of the endonuclease gene and its site [62, 65]. In the *delitto perfetto*-DSB system, the induction of the DSB stimulates oligo targeting more than 1,000-fold [59, 62]. A DSB can also stimulate oligo-mediated gene correction in mammalian cells [66, 67]. Considering the high efficiency of recombination

driven by oligos in the presence of a DSB, it is possible to exploit this phenomenon, not only in gene targeting procedures, but also to study mechanisms of break repair using oligos with modified sequences of choice.

1.5.3. RNA-templated DSB repair

Previously, the capacity to repair a DNA DSB using homologous sequences was only attributed to DNA molecules [68]. RNA can be copied to DNA through reverse transcription in retroviruses, retrotransposons, and at telomeres [69, 70]. Reverse-transcriptase (RT)-mediated events were observed in DSB repair, in which mRNA was converted to complementary DNA (cDNA) within yeast retrotransposon Ty1 particles and inserted at the break site of an HO endonuclease induced DSB at the mating type (MAT) locus [71, 72]. Such repair events do not involve homology between the RNA and the broken DNA, thus the repair of the broken DNA region containing an insertion could be mutagenic. Although several studies have reported direct interaction of RNA with chromosomal DNA, none of them involves DNA repair. An RNA transcript can be used as a primer by the *E. coli* replisome to continue leading-strand synthesis after a co-directional collision with RNA polymerase [73]. Artificial RNA templates injected into *Oxytricha*, a Ciliata Protozoa, were shown to mediate correct and precise DNA rearrangements, suggesting that RNA molecules can guide genome rearrangements [74]. Moreover, RNA can invade a DNA duplex and pair with the complementary DNA to form a three-strand R-loop structure both *in vitro* [75] and *in vivo* [76].

Synthetic RNA-only and RNA-containing oligos were found to serve as templates for DNA synthesis during the repair of a chromosomal DSB in yeast [77] (**Figure 1.5**). Much evidence was provided in this study that RNA can function as a direct template for DNA synthesis in a homology driven manner during DSB repair. The repair frequency declined with the increased size of the RNA tract when the length of the RNA tract was larger than 4 bp, suggesting that cDNA generated by Ty reverse transcriptase was unlikely to be a source of DSB repair. The deletion of the *SPT3* gene, which is essential for transposition and transcription of Ty1 and Ty2 elements, or the deletion of telomerase genes, *EST1* or *EST2*, did not affect the DSB repair events mediated by a homologous RNA template [78, 79]. Another evidence that RNA serves as a direct template is based on the fact that RNA-containing oligos have the same strand-bias targeting as their corresponding DNA-only oligos [77]. The strand-bias targeting was first identified by using synthetic DNA oligos [80] to correct a genetic defect in a reporter gene positioned downstream from a DSB site (**Figure 1.6**). Following the formation of the DSB, the 5' ends of the break are resected, exposing long 3' single-strand DNA ends. The synthetic DNA oligo complementary to the 3' end can anneal with the exposed ssDNA end on the chromosome, while the oligo complementary to the 5' end of the break cannot anneal with the 5' end if this has been degraded, and thus, it cannot serve as template to correct the defect in the reporter gene. Therefore, the oligo complementary to the 3' end of the break is more efficient at gene correction than the oligo complementary to the 5' end of the break. If the RNA-containing oligos are reverse transcribed to complementary single-strand DNA, the bias for these oligos would be opposite to the bias observed for the corresponding DNA-only oligos. If the RNA-containing oligos are reverse transcribed to

double strand cDNA, there would be no strand bias because the oligos would be double stranded. In the experiments presented in **Figure 1.6**, the RNA-containing oligos presented the same strand bias as the corresponding DNA-only oligos, thus excluding the possibility of reverse transcription [77]. The overall findings of this study established that a direct homologous exchange of genetic information can occur between RNA and DNA in yeast cells.

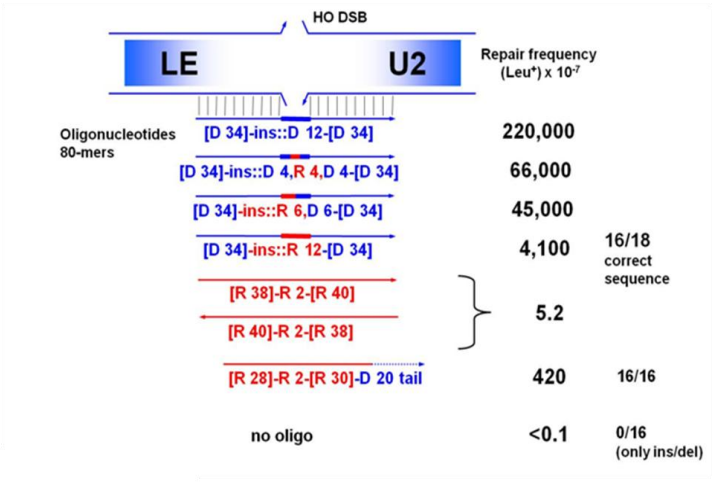


Figure 1.5 Repair of a DSB by RNA-containing oligos (modified from [77]). A DSB is induced in the *LEU2* gene on the yeast chromosome. RNA-containing oligos, RNA-only oligos, and their corresponding DNA oligo are transformed to repair the DSB. DNA is shown in blue; RNA is shown in red. Repair frequency and sequencing results of the repaired *LEU2* gene are shown on the right.

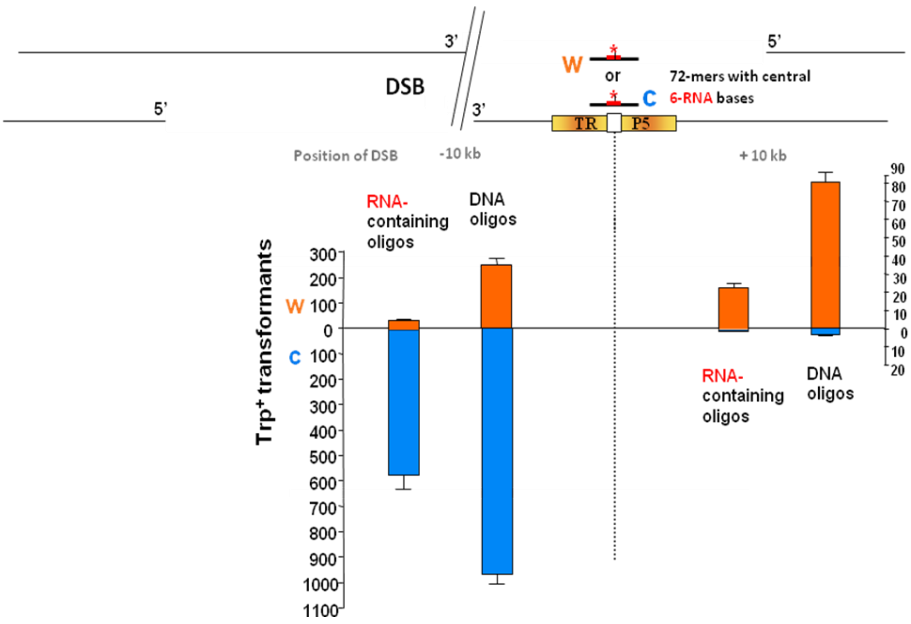


Figure 1.6 Strand bias targeting of RNA-containing oligos and their complementary DNA-only oligos (modified from [77]). W: oligo complementary to the 3' end; C: oligo complementary to the 5' end.

RNA-only and RNA-containing oligos have relatively low repair efficiency compared with their corresponding DNA oligos [77]. Two possible explanations are that (1) a single strand RNA is readily degraded by exonucleases [81, 82], (2) an RNA tract bound to DNA in an RNA/DNA hybrid can be easily recognized by RNase H1 and/or RNase H2 [20], which then cleave the RNA tract and prevent the repair. Studying the specificity of RNases H1 and RNases H2 *in vivo* will help to decipher the mechanism of RNA-mediated DNA repair.

1.6. RESEARCH GOALS

1.6.1. To explore the phenomenon of RNA-mediated DNA repair and modification in bacteria and human cells.

Our previous studies have demonstrated that RNA can serve as a template for DSB repair in the yeast *S. cerevisiae* using synthetic RNA-containing oligos. Here, we hypothesize that RNA-mediated DNA repair and modification is conserved in both prokaryotic and eukaryotic cells. We have developed an assay in the bacterium *E. coli* to correct a defective *lacZ* marker gene using RNA-containing oligos as a template. We have also utilized RNA-containing oligos to repair a DNA DSB in the green fluorescent protein (GFP) gene located either on a plasmid or in the genome in human embryonic kidney (HEK-293) cells.

1.6.2. To develop a molecular approach to generate RNA/DNA hybrids in genomic DNA.

RNA/DNA hybrids can form during DNA replication, repair and transcription. However, very little is known about the stability of RNA/DNA hybrids in cells and to which extent RNA/DNA hybrids can affect the genome integrity. Therefore, it is important to develop a molecular approach to generate RNA/DNA hybrids *in vivo*. We have set up a procedure for the incorporation of ribonucleotides into the genome of cells, starting with the eukaryotic model system yeast *S. cerevisiae*. This procedure to generate RNA/DNA hybrids at the chromosomal level can also be extended to other cell systems.

1.6.3. To identify the protein factors and DNA repair mechanisms that can target mispaired rNMPs embedded in DNA.

Numerous studies support the hypothesis that rNMPs are abundant in genomic DNA. We propose that the rNMPs in genomic DNA can form RNA/DNA mispairs, which could be a target for RNases H and DNA repair mechanisms, such as the DNA mismatch repair (MMR) system. To understand to what extent rNMPs may alter genome integrity and what factors affect their stability, we have utilized oligo-driven gene correction assays in the bacterium *E. coli* and the yeast *S. cerevisiae* and we have examined gene correction efficiency in different mutant backgrounds for these cell types.

CHAPTER 2

RNA-driven genetic changes in bacteria and in human cells

The study in Chapter 2 was published on Mutat. Res. 2011 [Epub ahead of print]

Shen Y.[1†], Nandi P.[1†], Taylor M.B.[1†], Stuckey S.[1], Bhadsavle H.P.[1], Weiss B.[2], Storici F.[1]

[1]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230;

[2]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322.

[†]Equal contribution to this work.

*The experiments in human cells were done by Nandi P., Taylor M.B., Stuckey S., and Bhadsavle H.P.

2.1. ABSTRACT

As recently demonstrated in the yeast *Saccharomyces cerevisiae* model organism using synthetic RNA-containing oligonucleotides (oligos), RNA can serve as a template for DNA synthesis at the chromosomal level during the process of double-strand break (DSB) repair. Herein we show that the phenomenon of RNA-mediated DNA modification and repair is not limited to yeast cells. A tract of six ribonucleotides embedded in single-strand DNA oligos corresponding to either lagging or leading strand sequences could serve as a template to correct a defective *lacZ* marker gene in the chromosome of the bacterium *Escherichia coli*. In order to test the capacity of RNA to modify DNA in mammalian cells, we utilized DNA oligos containing an embedded tract of six ribonucleotides, as well as oligos mostly made of RNA. These oligos were designed to repair a chromosomal break generated within a copy of the green fluorescent protein (GFP) gene randomly integrated into the genome of human HEK-293 cells. We show that these RNA-containing oligos can serve as templates to repair a DSB in human cells and can introduce base changes into genomic or plasmid DNA. In both *E. coli* and human cells, the strand bias of chromosomal gene correction by the single-strand RNA-containing oligos was the same as that obtained for the corresponding DNA molecules. Therefore, the RNA-containing oligos are not converted into a cDNA before annealing with complementary DNA. Overall, we demonstrate that in both bacterial and human cells, as in yeast, RNA sequences can have a direct role in DNA genetic modification and remodeling.

2.2. INTRODUCTION

In the central dogma of molecular biology, information flows from DNA to RNA. Information can be transferred from RNA to DNA by reverse transcription only in the special cases of retroviruses, retrotransposons and telomerase [83, 84]. Could there be functions for RNA-driven DNA modifications of which we are unaware? There might be profound implications in molecular biology if transfer of genetic information from RNA to DNA could be shown to be more general in nature.

Among the many roles ascribed to RNA in cells, we recently revealed that RNA can function as a template for double-strand break (DSB) repair and directly transfer information to chromosomal DNA in a model organism, the yeast *Saccharomyces cerevisiae* [25, 77]. Previously, the capacity to repair a DNA DSB by a homologous template molecule was attributed to pure DNA molecules [68], and none of the known DNA repair pathways had ever been shown to utilize RNA as a template. Reverse-transcriptase (RT)-mediated events were observed in DSB repair in yeast strains, in which mRNA was copied into DNA (cDNA) and inserted at an HO endonuclease-induced DSB at the mating type (MAT) locus [85, 86], or was used as a template for gene conversion with homologous genomic sequences [87, 88]. While yeast RT from the yeast transposon Ty is confined in the cytoplasm inside Ty particles [89], in human cells it was found that retrotranscription can be primed in the nucleus by the 3' end of a chromosomal break that captures the poly-A tail of a LINE1 retrotransposon RNA [90]. However, such events of break repair via LINE1 RNA did not involve homology between the RNA and the broken DNA. As reviewed by Storici [91], there have been prior speculations,

although without evidence, that RNA can be used as template in DNA repair or to directly transfer genetic information to genomic DNA. Differently, our data showed that RNA-containing oligonucleotides (oligos) with homology to a broken chromosome can efficiently repair a DSB under conditions where the previously described cDNA-dependent RNA repair processes were inactivated [77]. While there are several studies reporting the direct interaction of RNA with genomic DNA, none involve DNA repair or homologous targeting. In fact, RNA can invade a DNA duplex and pair with a complementary DNA region not only *in vitro* [92] but also *in vivo* [93], forming three strand R-loop structures, which have been regarded as sites of DNA replication in bacterial and mitochondrial DNA [94, 95]. RNA-guided DNA alterations were shown in *Tetrahymena* [96] and *Paramecium* [97]. Moreover, artificial RNA templates injected into the protozoa *Oxytricha trifallax* were recently shown to mediate correct and precise DNA rearrangements, suggesting that RNA molecules can guide genome modifications [96], thus supporting our results in yeast.

To further exploit the use of RNA-containing oligos, we have set up a procedure to generate desired RNA/DNA hybrids at the chromosomal level *in vivo* starting with the yeast system [98]. We have thereafter established that, in yeast *S. cerevisiae* and bacteria *E. coli* cells, mispaired ribonucleotides embedded in chromosomal DNA are sources of genetic modification (Shen *et al*., in press). In the present study, we demonstrate that the capacity of RNA to transfer genetic information to DNA is conserved from a prokaryotic to higher eukaryotic cell system, *i.e*. human cells. We also exclude the possibility that RNA-driven modification might be mediated by a cDNA reverse transcript. The results

we present, deviating from the central dogma of molecular biology, shed light on the capacity of RNA to play an active role in DNA editing and remodeling, which could be the basis of a wholly unexplored process of RNA-driven DNA evolution.

2.3. MATERIALS AND METHODS

2.3.1. *E. coli* strain and procedures.

The *E. coli* strain used in this study is BW1892 (*mutS*::Gm λ*c*I857Δ(*cro-bio*) *lacZ* [ΔGG (nt 1370-71); nt 1384 G→A]) [99], which has an AAG (Lac^-) codon in place of the GAG (Lac^+) codon of the wild type strain that differs from the GAA (glutamate) codon in the official sequence. The GG deletion prevents spontaneous reversion. Lac^+ transformants will have acquired both the GG segment and the GAG codon from the transforming oligos. The *E. coli* strain BW1892 has a *mutS* mutation and is thermoinducible for the beta protein of phage lambda, which makes it transformable by oligos [100]. LB media was used for routine growth. Lac^+ recombinants were selected on 1.5% agar plates containing Vogel-Bonner Medium E supplemented with lactose (0.1%), thiamine (1 μg/ml), and biotin (0.8 μg/ml) [99]. Transformations with single-stranded oligos were performed as previously described [100] using 100 ng of oligo per 50 μl of electrocompetent cells.

2.3.2. Human cell lines, plasmids and procedures.

Human embryonic kidney (HEK-293) cells were grown in Dulbecco's modified Eagle's medium, DMEM (Mediatech, Inc. Manassas, VA), supplemented with 10% heat-inactivated fetal bovine serum (Gemini, Bio-Products, West Sacramento, CA) and 1X Penicillin/Streptomycin (Lonza, Walkersville, MD). Cells were grown at 37˚C in a 5% CO_2 humidified incubator. 658D cells are HEK-293 derivative cells containing a randomly integrated copy of the GFP gene disrupted by a 35-bp region containing a stop

codon and the cutting site for the I-*Sce*I endonuclease [101]. Plasmid pA568 contains the GFP gene disrupted by the same 35-bp region containing the site for the I-*Sce*I endonuclease and is expressed under the CMV/CBA promoter [101]. Plasmid pA961 contains the I-*Sce*I endonuclease gene expressed under the CMV/CBA promoter [101]. Cells were transfected using polyethylenimine (PEI, Polysciences, Warrington, PA) transfection reagent in 24-well plates at a density of ~50,000 cells per well [102, 103]. In all transfection experiments in human cells, plasmid DNA was used in the amount of 0.5 μg while oligo DNA was 1.5 μg, unless otherwise indicated. In these experiments, the oligos and the plasmid were diluted in DMEM without supplements, vortexed in the presence of PEI, and then added to the wells 10 min later. Under these conditions, transfection of 0.5 μg of GFP expressing plasmid resulted in 80.5% GFP positive HEK-293 cells and 76% GFP positive 658D cells. Green fluorescent cells were visualized by fluorescent microscopy using a Zeiss Observer A1 microscope and an AxioCam MRm camera (Zeiss, Thornwood, NY). Frequencies of GFP positive cells were obtained by flow cytometric analysis using the BD LSR II Flow Cytometer (BD Biosciences, Sparks, MD). Either 50,000 or 100,000 cells were counted for each sample. Sorting of GFP positive cells was performed using the BD FACS Aria Cell Sorter (BD Biosciences).

2.3.3. Oligos and oligo alkali treatment.

DNA oligos were synthesized by Invitrogen (Carlsbad, CA) and Alpha DNA (Montreal, Quebec, Canada), while RNA-containing oligos were synthesized by Thermo Fisher Scientific (Dharmacon brand, Lafayette, CO) (**Table 2.1**). For experiments in *E. coli* cells, to establish the presence of ribonucleotides by using alkaline sensitivity, oligos at 5

μg/μl were incubated for 1 h at 65°C in water (no treatment) or in 0.5 M NaOH followed by neutralization with HCl 1.2 M and Tris-HCl 1 M pH 7.4. A high concentration of oligo was used to enable sufficient subsequent dilution of salt before electroporation. Transformations of bacterial cells were completed as described above with 100 ng of oligo. For experiments in human cells, 4 nmoles of oligos were treated with 1 M NaOH at 65° C for 1 h, and then neutralized with HCl 1.2 M and Tris-HCl 1 M pH 7.4. Alternatively, 4 nmoles of the same oligos were resuspended in water and Tris-HCl 1 M pH 7.4. 1.5 μg of oligos were used in the transfection experiments as described above.

2.3.4. Standard genetic and molecular biology techniques.

Standard genetics and molecular biology analyses were performed as previously described [77, 99]. PCR from the *E. coli* genome was performed by introducing single colonies directly into the PCR mix for the *Taq* Roche DNA polymerase (Basel, Switzerland) and using primers *LacZ1201F*: 5'-ACCGCTACGGCCTGTATGTG and *LacZ1703R*: GACGAAGCCGCCCTGTAAAC). Purified PCR products were submitted for sequencing to Eurofins MWG Operon (Huntsville, AL). To verify the sequence of the GFP DNA targeted by the DNA and the RNA-containing oligos in the human cells, green fluorescent sorted cells were utilized for genomic DNA extraction according to the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). The extracted genomic DNA was used as a template to amplify the sequence of the GFP marker using primers GFP.1 (5'-atggtgagcaagggcgaggagctg) and GFP.2 (5'- ttacttgtacagctcgtccatgcc) using iProof High Fidelity DNA Polymerase (BioRad, Hercules, CA). The PCR products obtained were run in an agarose gel to verify the correct size and were then cloned using the Zero Blunt

TOPO PCR Cloning Kit (Invitrogen) according to the instructions. The GFP sequence was then determined by sequence analysis using the standard primers for the SP6 promoter and the T7 promoter as indicated in the Zero Blunt TOPO PCR Cloning Kit protocol.

Table 2.1 DNA sequences of the oligos used in this study and of part of the GFP locus disrupted by the I-*Sce*I site

Name	Size	Structure from 5'	Sequence
LacZ.D	65	[D 24] ins::D 2 [D 12] mD 1 [D 26]	5'-GTAATCACCCGAGTGTGATCATCT**GG**TCGCTGGGGAAT**G**AGTCAGGCCACGGCGCTAATCACGAC
LacZ.D comp	65	[D 26] mD 1[D 12] ins::D 2 [D 24]	5'-GTCGTGATTAGCGCCGTGGCCTGACT**C**ATTCCCCAGCGA**CC**AGATGATCACACTCGGGTGATTAC
LacZ.R6	65	[D 22 R 2] ins::R 2 [R 2 D 10] mD 1 [D 26]	5'-GTAATCACCCGAGTGTGATCATCU**rGrG**UCGCTGGGGAAT**G**AGTCAGGCCACGGCGCTAATCACGAC
LacZ.R6 comp	65	[D 26] mD1 [D 10 R 2] ins::R 2 [R 2 D 22]	5'-GTCGTGATTAGCGCCGTGGCCTGACT**C**ATTCCCCAGCGA**rCrC**AGATGATCACACTCGGGTGATTAC
GFP80.D	80	[D 39] mD 1 [D 40]	5'-GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC**G**CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC
GFP80.D comp	80	[D 40] mD 1 [D 39]	5'-GGTTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGC**C**GTCTTGTAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGC
GFP.D	66	[D 32] mD 1 [D 33]	5'-ATCTTCTTCAAGGACGACGGCAACTACAAGAC**G**CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
GFP.D comp	66	[D 33] mD 1 [D 32]	5'-CAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGC**C**GTCTTGTAGTTGCCGTCGTCCTTGAAGAAGAT
GFP.R6	66	[D 30 R 2] mR 1 [R3 D 33]	5'-ATCTTCTTCAAGGACGACGGCAACTACAAGAC**rG**CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
GFP.R6 comp	66	[D 30 R 3] mR 1 [R 2 D 32]	5'-CAGGGTGTCGCCCTCGAACTTCACCTCGGCGCG**rC**GUCTTGTAGTTGCCGTCGTCCTTGAAGAAGAT
DT_GFP.R	80	[R 29] mR 1 [R 30] D 20 tail	5'-UUCUUCAAGGACGACGGCAACUACAAGAC**rG**CGCGCCGAGGUGAAGUUCGAGGGCGACACC-CCAGCGGAGTCCTTGGCCTA
GFP80.R	80	[R 39] mR 1 [R 40]	5'-GCGCACCAUCUUCUUCAAGGACGACGGCAACUACAAGAC**rG**CGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACC
GFP::I-*Sce*I site		5'---ATCTTCTTCAAGGACGACGGCAACTACAAGACC-TAAGCTCTCGAGATTACCCTGTTATCCCTAAGCTT-CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG---	

The structure of the oligos used in this study is described from the 5'end, with DNA sequences (D) shown in blue and RNA sequences (R) in red. Bases with homology to the chromosomal DNA are shown in brackets and underlined. Insertions are indicated as "ins::". Mismatches are indicated as "m". The sequence of the oligos is shown from the 5' end; the sequence changes introduced by the oligos are bolded and shown in blue if these consist of DNA bases and in red, with a lower case "r" on the left side if these consist of RNA bases. The *Bss*HII restriction site is underlined in the sequence of the GFP targeting oligos. The DNA oligos were desalted. The RNA–containing oligos were desalted and deprotected.

The tract of the GFP sequence disrupted by a 35-bp region containing the 18-bp I-*Sce*I cutting site is shown at the bottom. The I-*Sce*I site is underlined.

2.4. RESULTS AND DISCUSSION

2.4.1. RNA-driven gene correction in *E. coli*.

The capacity of RNA to modify genomic DNA is not limited to DSB repair. From experiments we performed, both in *S. cerevisiae* and in *E. coli*, we have found that RNA can serve as a template for genomic DNA modifications in circumstances where a DSB was not induced [77, 91, 98] (Shen *et al*, in press). When we first demonstrated RNA-driven DNA modifications in yeast cells we excluded the possibility that RNA was copied by a reverse transcriptase enzyme and converted into a cDNA intermediate before serving as template for DNA synthesis at the chromosomal level [77]. We exploited the finding that a DSB can activate strong strand-biased targeting by ssDNA oligos with homology to loci distant from the DSB site [80]. If a cDNA intermediate were formed from single-stranded RNA-containing oligos, the observed bias should have been opposite to that with the corresponding single-stranded DNA oligos, and there would have been no bias if double-stranded DNA were formed. As we previously showed [77], targeting with RNA-containing oligos was biased similarly to DNA-only oligos. Moreover, after eliminating the major sources of RT in yeast from Ty1 and Ty2 or telomerase by deleting the *SPT3* gene or the *EST1* and *EST2* genes, respectively, the frequency of RNA-templated DNA repair by multiple RNA-containing oligos remained similar to that obtained in wild-type cells in yeast [77]. Therefore, our data clearly demonstrated that the RNA-containing oligos can directly transfer genetic information to chromosomal DNA.

Unlike yeast cells that express the RT from the Ty transposons and from telomerase, *E. coli* K-12 and derivative strains do not have any RT [104] that could reverse transcribe the RNA-containing oligos into cDNA. Although all of our *E. coli* strains derive from the K-12 strain, to completely exclude the possibility that the RNA-containing oligos could be copied into DNA before interacting with the *E. coli* chromosomal DNA, we took advantage of the fact that gene correction by single-strand oligos in *E. coli* is known to be strongly affected by the direction of DNA replication. Specifically, in the λ recombination system, single-strand DNA oligos corresponding to lagging strand sequences are 10 to 100-fold more efficient at gene targeting than the complementary single-strand oligos corresponding to the leading strand [105]. Because of the strong strand bias for transformation, it is believed that the transforming oligos bind to the chromosome at the replication fork, in the gaps between the Okazaki fragments, and are subsequently joined to the lagging strand, thereby becoming templates for gene correction in the next round of DNA replication [105]. We hypothesized that if the RNA-containing oligos were interacting directly with chromosomal DNA these should display the same strand bias targeting as the corresponding DNA oligos. We, therefore, investigated the capacity of lagging and leading-strand oligos containing an RNA tract to transfer information to chromosomal DNA in bacterial cells, utilizing the *E. coli* strain BW1892 (see section 2.3.1). Strain BW1892 contains a defective *lacZ* gene in the bacterial genome. The *lacZ* mutant gene carries a missense mutation (GAG in place of AAG), and an additional 2-nucleotide (nt) deletion (deletion of GG) upstream from the substitution prevents spontaneous reversion to Lac^+ [99]. The BW1892 strain also contains a partially deleted λ prophage that expresses the Beta and exonuclease

recombination proteins of λ under the thermoinducible control of the λ*c*I*857* repressor. The Beta protein enables the efficient incorporation of electroporated oligos into DNA [100]. The *mutS* mutation in the host enhances the frequency of transformation by blocking mismatch repair, which would otherwise impede transfer of information from the oligos to the chromosomal DNA [100]. This strain can be efficiently transformed to Lac^+ with appropriate oligos that substitute the mutant G into an A and reintroduce the missing GG bases (**Figure 2.1**). Two RNA-containing oligos were used to correct these mutations. One brought back in the missing GG bases present within a 6-base RNA tract embedded in the DNA sequence of a 65-nt single-strand molecule also containing the correct 'A' to replace the 'G' of the substitution mutation (*LacZ.R6*); the second was the complementary oligo (*LacZ.R6 comp*) (**Table 2.1 and Figure 2.1**). These 6-RNA base-containing oligos and the corresponding DNA-only molecules (*LacZ.D* and *LacZ.D comp*) were transformed into the *E. coli* BW1892 strain. If the information is correctly transferred from the oligo to the chromosome, bacteria cells become Lac^+. Transformant colonies were obtained on the selective media. Results are shown in **Table 2.2**. The DNA-only oligo of the lagging strand could correct the defective *lacZ* gene 12.2-fold more efficiently than the complementary DNA-only oligo of the leading strand. In the BW1892 background, the RNA-containing oligo *LacZ.R6* of the lagging strand could serve as a template for gene correction. Importantly, gene correction by the *LacZ.R6* oligo was over two orders of magnitude (286-fold) more efficient than the *LacZ.R6 comp* oligo of the leading strand, revealing a strong strand bias in favor of the RNA-containing oligo of the lagging strand (**Table 2.2**). Barring an unknown mechanism for reverse transcription biased towards very specific sequences in *E. coli* K12, such as the sequence

of the *LacZ.R6* oligo, our data demonstrate that the RNA-containing oligos are not copied into cDNA before interacting with chromosomal DNA in *E. coli*, and can thus be used as templates for DNA synthesis directly at the chromosomal level.

The ability of both the RNA-containing oligos of the lagging and leading strands to transfer information to chromosomal DNA in *E. coli* cells was clearly not due to contamination with the DNA oligos. Pretreatment with alkali (NaOH) before transformation with oligos almost abolished gene correction by the RNA-containing oligos (**Table 2.3**). Therefore, the number of Lac^+ transformants produced by the *LacZ.R6* oligo was much greater than what could be explained by DNA contamination (which could be no more than 0.63 Lac^+ cells/10^7 cells). Five Lac^+ transformants obtained with the RNA-containing oligo *LacZ.R6* and five obtained with oligo *LacZ.R6 comp* were sequenced in the *lacZ* region. The chromosomal sequences were identical to the oligo sequences in 5/5 samples targeted by the *LacZ.R6* oligo and in 4/5 samples targeted by the *LacZ.R6 comp* oligo (data not shown). One clone targeted by the *LacZ.R6 comp* oligo displayed two additional substitutions (one within the sequence derived from the RNA tract of the oligo and one within the sequence derived from the DNA tract of the oligo), probably due to errors during oligo synthesis [52]. Overall, these results demonstrate that *E. coli* cells can utilize RNA as a template for DNA synthesis at the chromosomal level. Thus, RNA-containing molecules can be a source of genetic changes in *E. coli*.

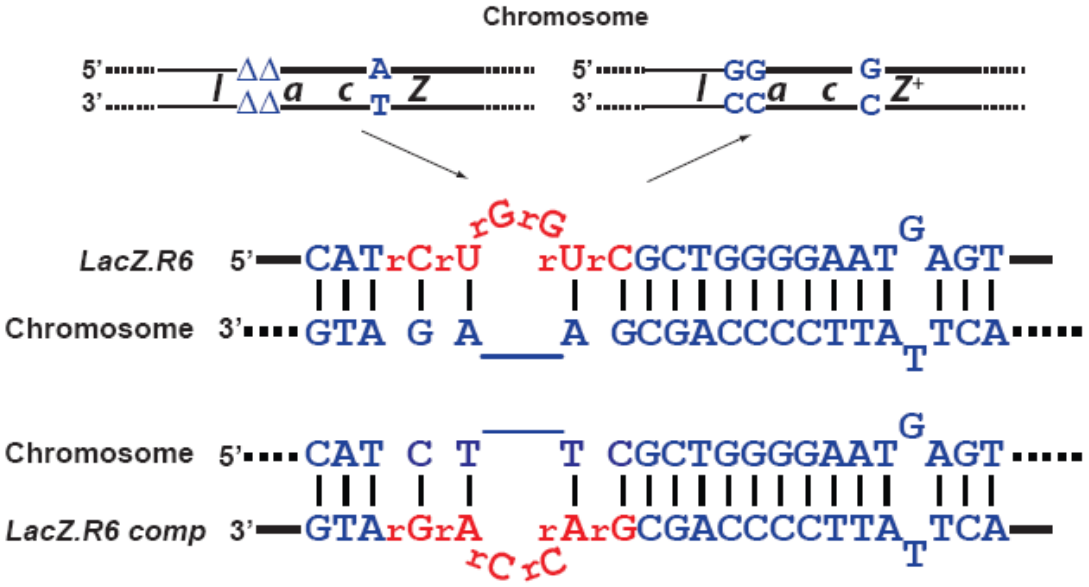


Figure 2.1 Scheme of the *E. coli* chromosomal region with the *lacZ* gene mutated and then repaired by the *LacZ.R6* oligo or the *LacZ.R6 comp* oligo. DNA bases are shown in blue and RNA bases in red with a small 'r' on the left.

Table 2.2 Strand-bias targeting of RNA-containing oligos in *E. coli*

Oligos	Lac$^+$ colonies		Strand†
LacZ.R6	189	(155-249)	Lagging
LacZ.R6 comp	0.66	(0.06-0.94)	Leading
LacZ.D	9,750	(8,060-12,400)	Lagging
LacZ.D comp	798	(364-1,480)	Leading
no oligo	<0.1	(0-0)	NA

Median and 95% confidence interval (in parentheses), or alternatively the range when the number of repeated experiments was <6, of *E.coli* Lac$^+$ transformant colonies reverting the two-base deletion mutation and one-base missense mutation per 10^7 viable cells. †The correspondence to the lagging or leading strand of DNA replication is indicated for each oligo.

Table 2.3 Alkali lability of RNA-containing oligos in *E. coli*

Oligos	Lac$^+$ colonies w/o NaOH		Lac$^+$ colonies w/ NaOH	
LacZ.R6	583	(394-772)	0.63	(0.31-0.63)
LacZ.R6 comp	0.47	(0-0.79)	<0.1	(0-0)
LacZ.D	15,300	(9,130-18,300)	4,100	(1,900-5,510)
LacZ.D comp	1,300	(1,220-1,850)	542	(441-611)
no oligo	<0.1		NA	

Median and range (in parentheses) of Lac$^+$ transformant colonies, per 10^7 viable cells, following treatment without NaOH or with NaOH. NA, not applicable.

2.4.2. RNA-mediated chromosomal repair in human cells.

We previously reported that the induction of a DSB stimulates up to 10,000-fold the gene correction frequency by both DNA and RNA-containing oligos in yeast cells [62, 77]. In mammalian cells, where homologous gene correction is very inefficient, a DSB was also shown to stimulate gene targeting [64, 101, 106]. More than 0.3% of human HEK-293 cells were corrected by DNA oligos when an I-*Sce*I DSB was induced at the target gene and the oligos had sequence corresponding to the sense strand of the target gene [107]. With the goal of detecting gene modification by RNA-containing oligos in mammalian cells, we exploited the break induction system developed by Porteus and Baltimore in HEK-293 derivative cells (658D cells) [101]. The 658D cell line contains a copy of the green fluorescent protein (GFP) gene randomly integrated into the genome of the HEK-293 cells and disrupted by a region containing the site for the I-*Sce*I endonuclease [101] (**Table 2.1**). A DSB can be induced in the middle of GFP by transfecting the 658D cells with a plasmid that expresses the I-*Sce*I endonuclease [101]. To test if RNA-containing oligos could be used to repair a DSB generated in the genome of human cells and transfer genetic information to genomic DNA, we designed several oligos containing RNA tracts of different lengths and with homology to both sides of the I-*Sce*I-break site in the GFP gene of the 658D cells (**Table 2.1**). These oligos were tested for their ability to mediate break repair and restore the functionality of the GFP marker gene (**Figure 2.2A and Table 2.1**). All oligos were also designed to contain a silent mutation in GFP to generate a *Bss*HII restriction site (**Table 2.1**). Initially, the RNA-containing oligo *GFP.R6* or the corresponding control DNA-only oligo *GFP.D* was cotransfected together with the I-*Sce*I endonuclease-expressing vector into the 658D cells. Remarkably, the RNA-containing

oligo *GFP.R6* generated green fluorescent cells following transfection in human cells (**Figure 2.2B**). The repair frequency of GFP by the different oligos was determined by flow cytometric analysis. As shown in **Figure 2.2C**, the *GFP.R6* oligo containing a tract of 6 ribonucleotides was able to restore the functionality of GFP and presented a repair frequency only 4.4-fold less than the corresponding DNA oligo *GFP.D*. To determine whether oligos with an RNA-only region homologous to the broken chromosomal GFP marker gene could also serve as a template for break repair and DNA modification, we designed oligo *DT_GFP.R* with 60 RNA bases homologous to GFP and a 3'-DNA non-homologous tail of 20 bases (**Table 2.1**). As previously demonstrated in our yeast study, an oligo containing 60 RNA bases with homology to broken yeast chromosomal ends and a 20-base non homologous DNA tail at the 3' end was about 100-fold more effective at DSB repair than a corresponding RNA-only molecule [77]. The DNA tail is designed to protect the RNA portion from degradation by nucleases and does not participate in the repair process. Following transfection of oligo *DT_GFP.R* in the 658D cells, we found that this oligo generated a significant number of green fluorescent human cells with a frequency of 1/100,000 cells (**Figure 2.2C**). In the absence of oligos, the frequency of GFP positive cells was 1/400,000 (**Figure 2.2C**).

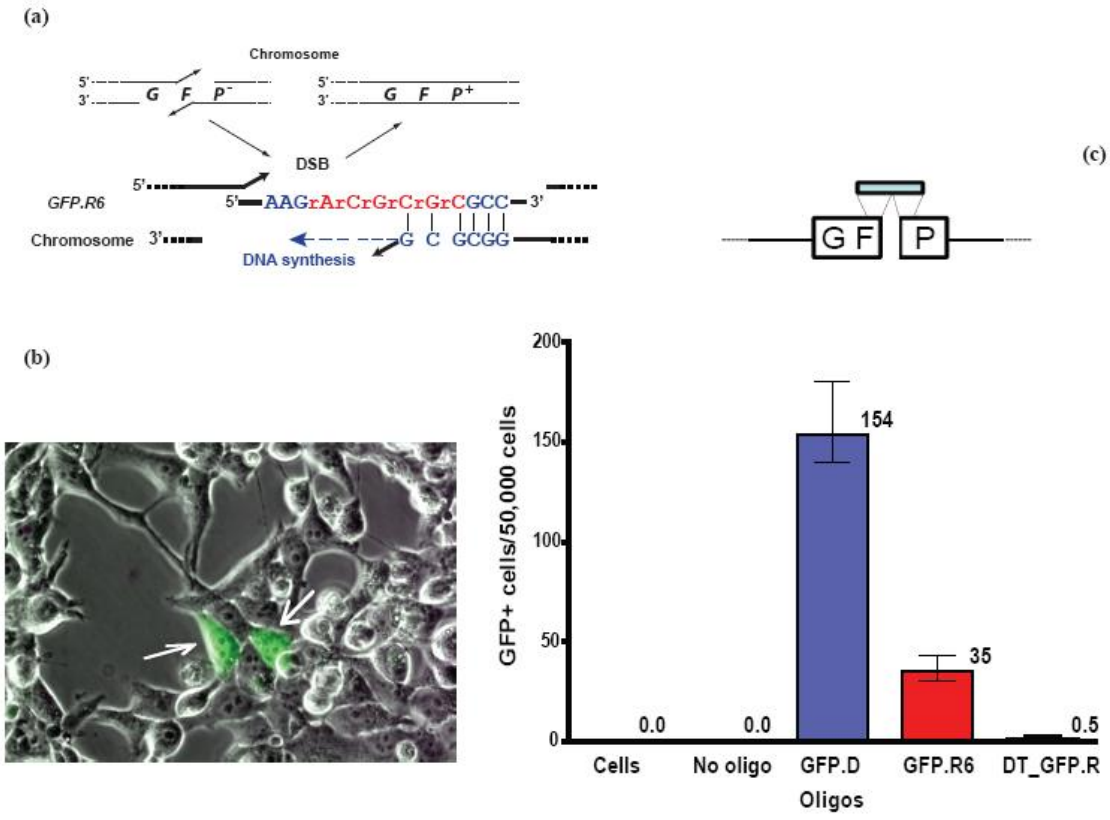


Figure 2.2 Chromosomal DNA modifications mediated by RNA-containing oligos in human cells. (a) Scheme of the human chromosomal region with the GFP gene cut by the I-*Sce*I endonuclease and then repaired by the *GFP.R6* oligo. DNA bases are shown in blue and RNA bases in red with a small 'r' on the left. (b) Micrograph at 40x of human green fluorescent cells (pointed to by the white arrows) targeted by the RNA-containing oligo *GFP.R6*. (c) Frequencies of human GFP^+ cells repaired by the DNA oligo (blue bar) or the RNA-containing oligos (red bars) per 50,000 transfected cells; median (value shown at the top right of each bar) and range are shown. While the median is 0.0, there were <1/400,000 GFP^+ cells for the 'Cells' sample and 1/250,000 in average for the 'No oligo' sample. Except for the 'Cells' sample, the 'No oligo' and the 'Oligos' samples were transfected together with the plasmid expressing the I-*Sce*I endonuclease. A sketch of chromosomal GFP repair by oligos is shown above the graph.

To exclude the possibility that DNA contamination in our GFP RNA-containing oligos was responsible for the transfer of genetic information, we treated the RNA-containing and DNA-only oligos with alkali prior to the experiment of gene correction at the chromosomal level (**Figure 2.3A**) or at the plasmid level (**Figure 2.3B**, as presented in section 2.4.4). Data presented in **Figure 2.3A** show that targeted repair of the chromosomal GFP was strongly reduced after alkali treatment of the RNA-containing oligos, thus excluding contamination with homologous DNA sequences in our RNA-containing oligo solutions. To confirm that correction of the GFP gene was due to oligo targeting and not by repair through end joining, green fluorescent cells obtained from *GFP.R6* and *GFP.D* oligos were sorted. After sorting, the population of green fluorescent cells reached ~ 75%, and 10^6 cells were prepared for genomic extraction. Genomic DNAs from the extracts were then used in PCR reactions to amplify the GFP gene (**Figure 2.4**). The PCR products were cloned using the Zero Blunt TOPO PCR Cloning Kit and digested with *Bss*HII or I-*Sce*I to verify the correction of the GFP gene. About 36% (20/55) of the TOPO clones analyzed and derived from sorted green fluorescent cells targeted by *GFP.R6* were still cut by I-*Sce*I, and ~10% (5/55) were cut by *Bss*HII. We sequenced 4 samples cut by *Bss*HII, and all 4 had the sequence carried by the RNA-containing oligo with no additional mutations. Of the remaining 54% of clones, 6 were sequenced, and none were shown to contain the GFP sequence (not shown). The presence of uncorrected GFP sequences can be explained by the fact that even after sorting of the green fluorescent cells, there were still several non-fluorescent cells present (~ 25%, not shown). Two clones cut by *Bss*HII derived from cells targeted by the DNA-only oligo *GFP.D* were sequenced and contained no other additional mutations beyond the expected

*Bss*HII site (data not shown). These data showed that an RNA-containing oligo, similarly to a corresponding DNA-only molecule, can serve as a template to repair the broken GFP gene and transfer information to human chromosomal DNA.

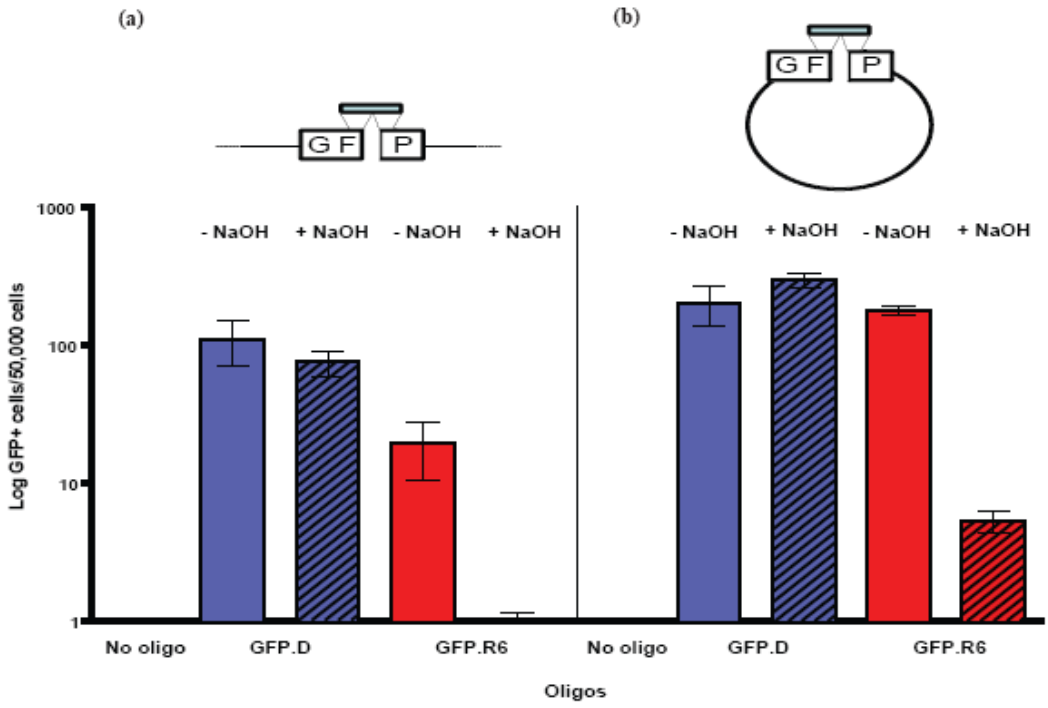


Figure 2.3 Alkali lability of RNA-containing oligos in human cells. Frequencies in Log scale of human GFP$^+$ cells repaired at the (a) chromosomal level or (b) the plasmid level by the DNA oligo following no treatment (-NaOH, blue bar) or treatment with alkali (+NaOH, blue bar with stripes), or by the RNA-containing oligo following no treatment (-NaOH, red bar) or treatment with alkali (+NaOH, red bar with stripes) per 50,000 transfected cells; median and range are shown. Sketches of chromosomal and plasmid GFP repair by oligos are shown above the graphs.

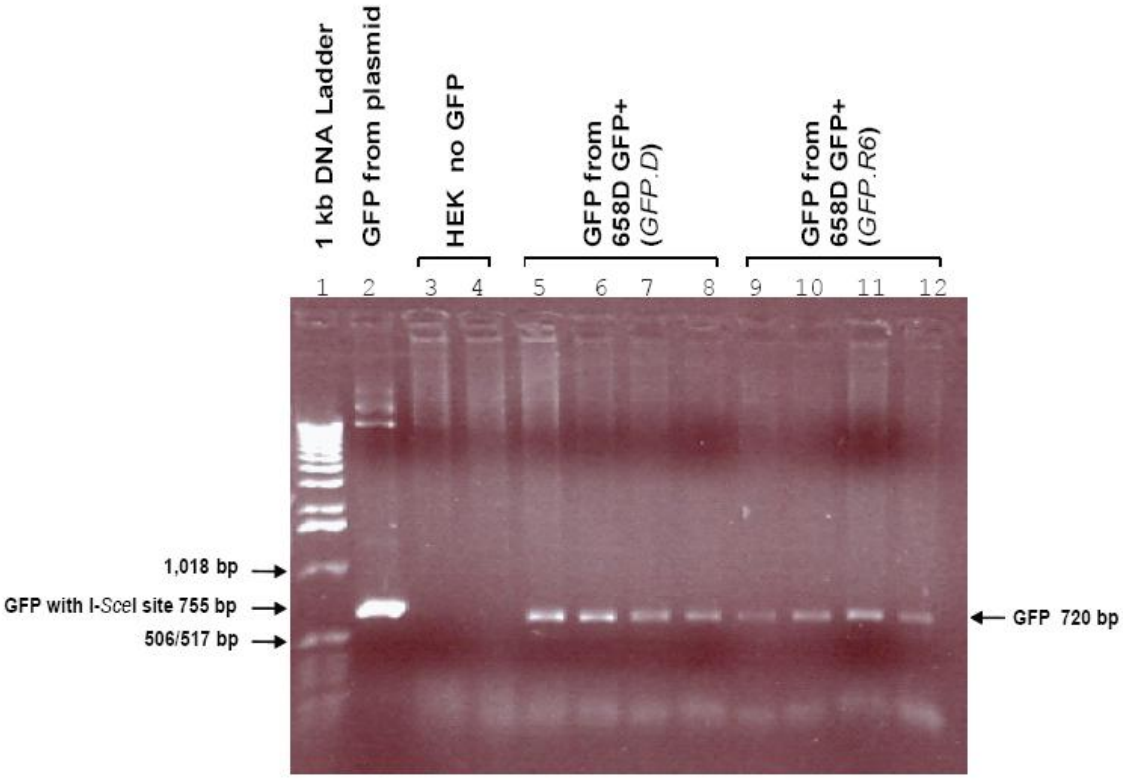


Figure 2.4 Detection of GFP bands following PCR from genomic DNA of human green fluorescent cells targeted by the DNA-only oligo *GFP.D* or the RNA-containing oligo *GFP.R6*. Lane 1, 1 kb DNA Ladder; lane 2, GFP band amplified from the pA658 vector containing a region with the I-*Sce*I site in the middle of GFP (755 bp); lanes 3 and 4, no PCR product detected from HEK-293 cells; lanes 5-8, GFP product (720 bp) from sorted green fluorescent 658D cells targeted by oligo *GFP.D*; lanes 9-12, GFP product (720 bp) from sorted green fluorescent 658D cells targeted by oligo *GFP.R6.*

2.4.3. RNA-containing oligos have the same strand bias as corresponding DNA oligos in repairing chromosomal GFP

To accomplish DSB repair on human chromosomal DNA and restore a functional GFP gene, the RNA tract of the RNA-containing oligos must be used as a template. The appearance of the sequence with the *Bss*HII site carried on the RNA tract in the genome of human green fluorescent cells suggests that the RNA-containing oligos are participating directly in the repair of GFP. Human cells contain abundant sources of reverse transcriptase activity, including LINE1 and L1 elements [90, 108]. Thus, considering the large quantity of RTs, it is not realistic to eliminate all RT sources present in the human cells. To show that the RNA-containing oligos can directly transfer information to chromosomal DNA in human cells without being first reverse transcribed into a cDNA molecule, we set up an experiment to determine if the RNA-containing oligos displayed a strand-biased repair of the chromosomal GFP. The DNA-only oligo spanning the sequence of the GFP sense strand (*GFP.D*) was always more efficient at repairing the chromosomal DSB in GFP than the complementary oligo comprising the GFP antisense sequence (*GFP.D comp*) (**Figure 2.5**). The bias did not depend on the quality of the oligos, as it reproduced even with different oligo preparations ordered from different companies (not shown). Moreover, the bias was observed only when the oligos were correcting the chromosomal broken GFP since no bias was revealed when the same oligos (*GFP.D* and *GFP.D comp*) were correcting the broken GFP on the plasmid (see below section 2.4.4). It remains unknown if the sense strand of the chromosomal GFP corresponds to the lagging strand, as the GFP locus was randomly integrated in the human genome and the site of integration was not determined [101]. In light of the

observed strand bias in the chromosomal targeting by single-stranded DNA oligos, we envisioned that if cDNA intermediates were formed following reverse transcription of the RNA-containing oligos, the observed bias of GFP repair by RNA-containing oligos would be opposite to that obtained with the corresponding DNA oligos, or there would be no bias if the DNA copies were double stranded. As shown in **Figure 2.5**, GFP targeted repair by the RNA-containing oligos displayed the same strand bias as the corresponding DNA-only molecules. Two clones cut by *Bss*HII derived from cells targeted by the DNA-only oligo *GFP.D comp* and three clones cut by *Bss*HII derived from cells targeted by *GFP.R6 comp* were sequenced and contained no other additional mutations beyond the expected *Bss*HII site (data not shown). Thus, it appears that the RNA-containing oligos are not copied by an RT function into cDNA before transferring the information to chromosomal DNA in human cells. We conclude that human cells have the ability to use RNA as a template within the chromosome.

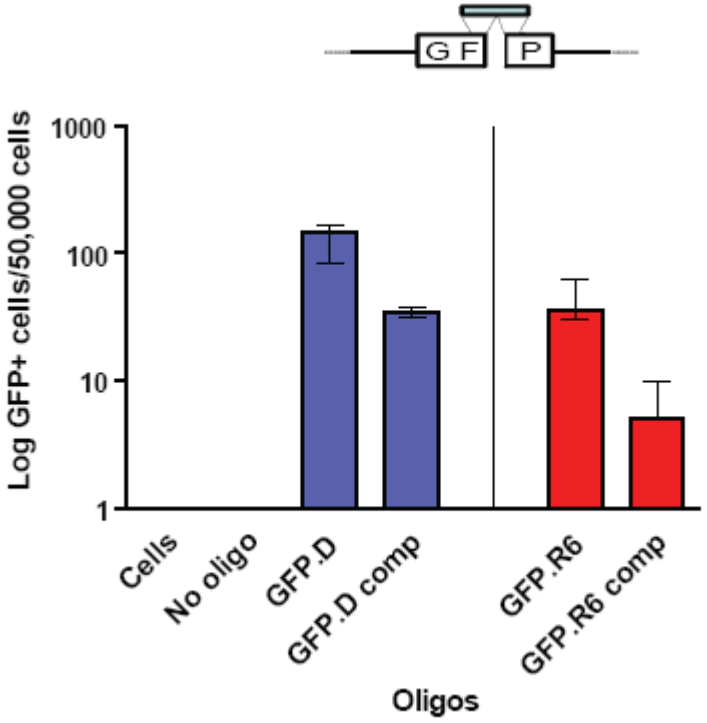


Figure 2.5 Strand-bias targeting to broken chromosomal GFP in human cells. Frequencies in Log scale of human GFP^+ cells repaired by the DNA oligos (blue bars) or the RNA-containing oligos (red bars) per 50,000 transfected cells; median and range are shown. Except for the 'Cells' sample, the 'No oligo' and the 'Oligos' samples were transfected together with the plasmid expressing the I-*Sce*I endonuclease. A sketch of chromosomal GFP repair by oligos is shown above the graph.

2.4.4. RNA-mediated plasmid repair in human cells

In order to test if the RNA-containing oligos were proficient at DSB repair also on extrachromosomal elements in human HEK-293 cells, we designed an experiment to examine the frequency of DSB repair on a plasmid. A plasmid construct (pA568) with the GFP gene interrupted by the I-*Sce*I site [101], the same gene sequence as the one integrated in the genome of the 658D cell line, was linearized *in vitro* by restriction digestion with the I-*Sce*I endonuclease (New England Biolabs, Ipswich, MA). The gapped vector was then cotransfected in the HEK-293 cells together with different RNA-containing oligos or control DNA oligos. As shown in **Figure 2.6**, frequencies of GFP repair obtained with oligos *GFP.R6* and *GFP.R6 comp* were about 10-fold less than those obtained with the corresponding DNA oligos (*GFP.D and GFP.D comp*). Although some strand bias was observed as well when the oligos repaired the broken GFP on the plasmid, it was not statistically significant. It is possible that the sequence context or the chromatin structure may differently affect chromosomal targeting by single-strand oligos from plasmid targeting. We also excluded the possibility of DNA contamination in our RNA-driven plasmid correction assay by demonstrating alkaline lability for oligo *GFP.R6* (**Figure 2.3B**).

The 80-mer *DT_GFP.R* has an RNA-only homologous region and a non-homologous 20-nt DNA flap at the 3' end. When it was cotransfected with gapped pA658, it could also serve as a template to repair the break and restore the correct sequence of the GFP gene in the HEK-293 cells, although it was ~ 100-fold less efficient than the DNA-only 80-mers *GFP80.D* and *GFP80.D comp* (**Figure 2.6**). We had to use 5-fold more oligo

GFP80.R (up to 10 μg) to obtain a detectable number of GFP^+ cells comparable to that obtained with oligo *DT_GFP.R*, which, in addition, contained an RNA stretch of homology to GFP shorter than the *GFP80.R* (**Figure 2.6 and Table 2.1**). It is likely that the DNA tail of oligo *DT_GFP.R* could in part prevent end-degradation of the RNA tract by exoribonucleases, which have a major role in RNA surveillance [109]. Overall, these data suggest that the RNA-containing oligos can transfer information not only to genomic DNA, but also to extrachromosomal DNA in the HEK-293 cells.

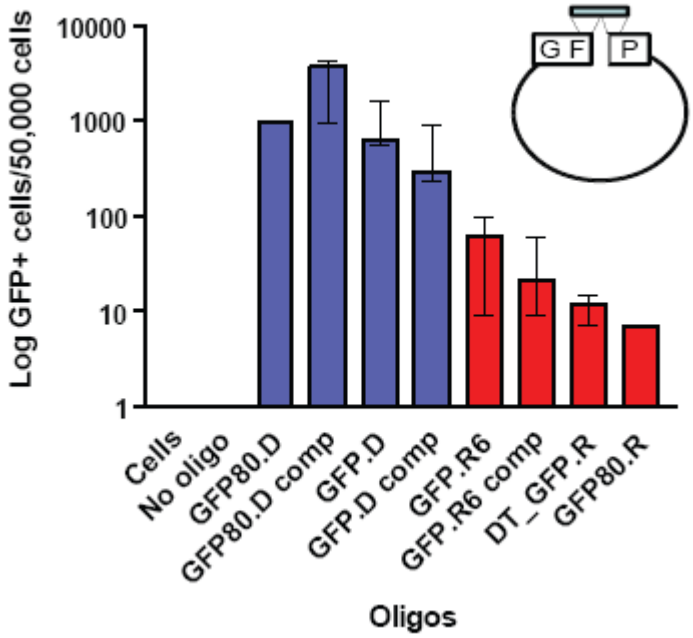


Figure 2.6 Plasmid DNA modifications mediated by RNA-containing oligos in human cells. Frequencies in Log scale of human GFP^+ cells repaired by the DNA oligos (blue bars) or the RNA-containing oligos (red bars) per 50,000 transfected cells; median and range are shown. Except for the 'Cells' sample, the 'No oligo' and the 'Oligos' samples were transfected together with the linearized pA568 plasmid. A sketch of plasmid GFP repair by oligos is shown above the graph.

2.5. CONCLUSIONS

Not long ago, it was considered that the major role played by RNA was to carry genetic information for protein translation. At present, we know that many RNAs are not used to make protein. Non-coding RNAs like small interfering RNAs, microRNAs and Piwi-interacting RNAs have emerged as fine control tools, which are important for the regulation of transcription and translation [110]. Although an astonishing variety of RNA functions have been found in the last few decades, it has always been very difficult to determine if any RNA has the capacity to genetically modify the DNA of cells. In this study we show that the ability of RNA to transfer genetic information to DNA is not limited to the yeast system in which the phenomenon was first discovered. By using RNA-containing oligos, we have found that RNA can function as a template for DNA synthesis without being reverse transcribed into cDNA, not only in yeast but also in *E. coli* and in the human HEK-293 cells. These findings establish that a direct flow of genetic information from RNA to DNA can occur in organisms as diverse as bacteria and humans, and thus, it can be a significant source of genetic variation.

How likely is the flow of information directly from RNA to genomic DNA without any specific reverse transcriptase function and without the requirement of any cDNA intermediate? Unexpectedly, as mentioned above, there are an increasing number of reports supporting the concept that RNA could guide genetic alterations. Moreover, there is a rising amount of evidence indicating that ribonucleotides can be incorporated into DNA during DNA synthesis and replication ([111] and references therein) and could thus

form hybrids with DNA. It is also possible that short RNA tracts are incorporated in DNA following primase inclusion of deoxyribonucleotides in addition to ribonucleotides in the primer sequence [8, 9]. It is worth noticing that ribonucleotides in DNA can distort the double helix [11, 112] and can form quite stable duplexes with DNA if enriched in ribonucleotide G [113]. It is however not known to what extent such RNA/DNA hybrids can be tolerated in DNA and to what level these hybrids can affect the genetic integrity of cells.

Interestingly, it has been postulated that the flow of information from RNA to DNA could lead to DNA recoding events in the nervous system that could be the basis for permanent storage of long term memories [114, 115]. Yet, the molecular mechanisms by which RNA could repair or edit DNA in neuronal cells are not known.

The goal of future research is to understand the mechanisms by which RNA can directly transfer information to the DNA of cells and to reveal the circumstances in which RNA information can flow to DNA.

2.6. ACKNOWLEDGMENTS

The authors wish to thank Patrick Ruff and Kyung Duk Koh for critical reading of the manuscript and helpful suggestions. The work was supported in part by the Georgia Cancer Coalition grant (R9028) and the NSF grant 1021763 to F. S.

CHAPTER 3

Generation of RNA/DNA hybrids in genomic DNA by transformation using RNA-containing oligonucleotides

The study in Chapter 3 was published on J. Vis. Exp. (45), 2010. Shen, Y. and Storici, F.

School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230;

3.1. ABSTRACT

Synthetic short nucleic acid polymers, oligonucleotides (oligos), are the most functional and widespread tools of molecular biology. Oligos can be produced to contain any desired DNA or RNA sequence and can be prepared to include a wide variety of base and sugar modifications. Moreover, oligos can be designed to mimic specific nucleic acid alterations and thus, can serve as important tools to investigate effects of DNA damage and mechanisms of repair. We found that Thermo Scientific Dharmacon RNA-containing oligos with a length between 50 and 80 nucleotides can be particularly suitable to study, *in vivo*, functions and consequences of chromosomal RNA/DNA hybrids and of ribonucleotides embedded into DNA. RNA/DNA hybrids can readily form during DNA replication, repair and transcription, however, very little is known about the stability of RNA/DNA hybrids in cells and to which extent these hybrids can affect the genetic integrity of cells. RNA-containing oligos, therefore, represent a perfect vector to introduce ribonucleotides into chromosomal DNA and generate RNA/DNA hybrids of chosen length and base composition. Here we present the protocol for the incorporation of ribonucleotides into the genome of the eukaryotic model system yeast *Saccharomyces cerevisiae*. Yet, our lab has utilized Thermo Scientific Dharmacon RNA-containing oligos to generate RNA/DNA hybrids at the chromosomal level in different cell systems, from bacteria to human cells.

3.2. PROTOCOL

3.2.1 Rationale

Exploiting the use of Thermo Scientific Dharmacon oligos, 50 to 80-mers, here we present a procedure, based on a gene correction assay, by which oligos can transfer genetic information to genomic DNA of yeast cells, following annealing to the target chromosomal DNA. Successful targeting by oligos is scored by the appearance of yeast colonies displaying the expected phenotype. If the desired genetic modification is carried within one or more ribonucleotides incorporated in the oligo sequence, the genetic information flows directly from the RNA tract to the chromosomal target DNA, thus, an RNA/DNA hybrid forms at the chromosomal level during the targeting process. The ribonucleotide/s of a Thermo Scientific Dharmacon RNA-containing oligo can serve as template for gene targeting via DNA repair synthesis or can be embedded into the DNA and serve as template during DNA replication. In these experiments we use the yeast *Saccharomyces cerevisiae* eukaryotic model system; however, a similar approach can also be applied in other organisms or cell types. In this video, we use a Thermo Scientific Dharmacon oligo designed with homology to a target yeast genomic locus and containing one ribonucleotide in the middle to correct a mutation in the target gene. After transformation with the RNA-containing oligo, we observe the correction of the targeted site at a certain frequency, which indicates that the RNA tract of the oligo could be incorporated into the chromosomal DNA and serve as template for DNA synthesis during DNA replication. The genetic information carried within the RNA tract is stably transmitted to the following cell generations. Here we describe the steps of yeast cell

transformation by the Thermo Scientific Dharmacon RNA-containing oligo and the approach to detect the RNA information transferring into chromosomal DNA.

3.2.2. Design of the RNA-containing Oligo used in this Experiment

We have designed an RNA-containing oligo to correct a genetic defect in the yeast *S. cerevisiae* chromosomal DNA at the *trp5* locus. The yeast strain used in the protocol contains a mutant *trp5* gene with a two-base deletion and one nonsense mutation (**Figure 3.1A**). Such yeast strain is a tryptophan auxotrophic mutant and does not form colonies on media without tryptophan. The DNA sequence of the TRP5 gene is available at the Saccharomyces Genome Database (http://www.yeastgenome.org/). The Thermo Scientific Dharmacon RNA-containing oligo used in this protocol is a 65-mer with a 2-base DNA insertion, designed to correct the deletion mutation and one ribonucleotide designed to correct the nonsense mutation of the *trp5* allele (**Figure 3.1A**). The sequence of the oligo is designed as follows:
5'-AAAAGGGTTTTGATGAAGCTGTCG-CG-GATCCCACATTCTG-rG-GAAGACTTCAAATCCTTGTATTCT. This oligo is synthesized by Thermo Scientific Dharmacon (Lafayette, CO) at 200 nM scale, and is desalted, deprotected and used without PAGE purification.

3.2.3. Preparation of the RNA-containing Oligo for Yeast Transformation (modified from Storici *et al*., 2007 [77])

1. Wipe materials that will be used for the experiment including oligo tubes, pipettes, vortex, racks, experimental area and gloves worn by the investigator with RNase decontamination solution to remove potential RNase contamination before

everything starts. Use RNase-free water, chemical reagents, tubes and pipette tips in all steps. Every step in this experiment should be RNase-free.

2. Resuspend the Thermo Scientific Dharmacon RNA-containing oligo received from the company to 250 pmoles/µl stock solution with RNase-free water and vortex vigorously to dissolve the pellet. Store at -80°C.
3. Immediately before the transformation, thaw the RNA-containing oligo on ice and dilute to 50 pmoles/µl with RNase-free water in RNase-free tubes. Each transformation requires 1 nmole of the RNA-containing oligos.
4. Denature a chosen amount of the RNA-containing oligos on the 100 °C heat block for 2 min to eliminate secondary structures of the oligos.
5. Immediately after denaturation place the tube on ice. Keep on ice till transformation.
6. As control in the experiment a corresponding DNA-only oligo is used. This oligo is thawed from -20 °C and prepared for the transformation as the RNA-containing oligo, as described above.

3.2.4. Transformation of Yeast Cells using the RNA-containing Oligo

1. Inoculate 5 ml of rich YPD liquid medium with the *trp5* mutant yeast cells and grow at 30 °C overnight (see Materials).
2. Transfer 1.5 ml of the overnight culture into 50 ml of YPD liquid medium.
3. Incubate cells in a 30 °C shaker (225 rpm) for 4h.
4. Prepare Solution 1 and Solution 2 immediately before transformation in RNase-free tubes (see Materials).

5. Transfer the cell culture to a 50 ml RNase-free tube and spin at 3000 rpm, corresponding to 1562 g, for 2 min. The pellet of the cell precipitation is approx. 0.3 cm^3.
6. Remove the supernatant and wash cells with 50 ml of RNase-free water and spin at 3000 rpm for 2 min.
7. Repeat step 6 for 5 times to get rid of the culture medium and RNases that could be present in the medium as much as possible.
8. Remove the supernatant and resuspend cells in 5 ml of Solution 1 and spin at 3000 rpm for 2 min.
9. Remove supernatant and resuspend cells in 250 µl of Solution 1. This amount of cells is sufficient for 7-8 transformations.
10. Aliquot 50 µl of the cell suspension in RNase-free microcentrifuge tubes, add 20 µl of RNA-containing oligo working solution (1 nmole), or 20 µl of DNA-only oligo working solution (1 nmole), or 20 µl of sterile water with no oligo for the negative control. Then add 300 µl of Solution 2 for each transformation reaction. There is no need to add salmon sperm DNA in the process of transformation, as the oligos act as carrier themselves.
11. Vortex vigorously to mix components homogenously.
12. Incubate transformation reactions at 30 °C for 30 min in a shaker.
13. Heat shock at 42 °C for 15 min.
14. Spin down cells at 5000 rpm, corresponding to 2236 g, for 4 min.
15. Remove supernatant and resuspend cells in 100 µl of sterile water.

16. Take an aliquot of this cell suspension and dilute it with sterile water by 100,000 fold and plate cells on one YPD plate using approx. 15 sterile glass beads and incubate at 30 ℃ for 2 days.
17. Plate all resuspended cells from each transformation reaction on one Petri dish of synthetic complete solid medium without tryptophan (SC-Trp) using approx. 15 sterile glass beads and incubate at 30 ℃ for 4-5 days (**Figure 3.2**).

3.2.5. Analysis of Gene Correction by the RNA-containing Oligo

1. Count the number of colonies grown on the selective medium (**Figure 3.2**) as well as on YPD medium to calculate the gene correction frequency for the RNA-containing oligo, the DNA-only oligo and for the no-oligo control. Compare the number obtained. Spontaneous reversion rate of the *trp5* alleles with the two-base deletions and the nonsense mutation is less than 10^{-9} in the used yeast strain, thus we expect no colonies formation on the selective medium when no oligos are added to the cells.
2. Streak out several randomly picked transformant colonies onto YPD medium to obtain single colony isolates. Wait two days for colony growth, then take several (at least 5) single colonies and make patches on YPD and on selective medium.
3. Design a pair of primers to amplify the region (250-1,000 bp) targeted by the RNA-containing oligo by colony PCR (**Figure 3.1B**). The procedure for colony PCR is as follows (modified from Storici and Resnick, 2006 [116]).
    1. Resuspend cells (approx. 1 mm^3) taken from the individual patches in 50 μl of water and add 1 unit of lyticase. Incubate at room temperature for 10

min, followed by incubation in a heat block at 100 °C for 5 min to break the cells and release genomic DNA in solution.

2. PCR conditions: The PCR reaction includes 10 µl of the cell resuspension solution, 50 pmoles each of forward and reverse primers, 1 µl of 10 mM dNTPs, 1 unit of Taq polymerase, 5 µl of 10x buffer and is adjusted with sterile water to a final volume of 50 µl. The PCR program is 3 min at 95 °C; 30 cycles of 30 s at 95°C, 30 s at 55 °C, and 1 min at 72 °C; a final extension time of 7 min at 72 °C; then samples are kept at 4 °C. An extension time of 1 min/kb is assumed for this reaction.
3. Following PCR, samples are run on a 1% or 2% (depending on the expected size of the PCR product) agarose gel for observation of the PCR product (**Figure 3.3**).

4. If the genetic information transferred by the RNA-containing oligo generates a new restriction site in the yeast genomic target region (**Figure 3.1B**), it is possible to verify the correct transfer of information by digesting the PCR product with the appropriate restriction enzyme. If no restriction site is generated by the RNA-containing oligo, go to step 6. Digest PCR products using a specific restriction enzyme. The digestion reaction includes 6 µl of PCR product, buffer, BSA (may not be needed for some enzymes, see instruction for the enzyme used), 0.5 µl of restriction enzyme, and sterile water to 15 µl. Samples are incubated for 1 h at the temperature specific for the enzyme used.

5. Run an undigested sample together with the digested samples on the same row of a 2% agarose gel to observe the genetic modification transferred by the RNA tract of the RNA-containing oligo (**Figure 3.3**).
6. Purify PCR products by using a PCR purification kit and prepare them for DNA sequencing. Submit samples for sequencing with the same primers used to amplify the product.
7. Analyze DNA sequencing results using software that allows alignment of multiple sequences with a chosen reference sequence (**Figure 3.4**).

3.2.6. Alkali Treatment of the RNA-containing oligo (Figure 3.5)

1. For each reaction, add 1 nmole (4 µl of 250 pmoles/µl stock solution) of the RNA-containing oligo, or DNA-oligo into a 1.5 ml tube.
2. Add 4 µl of 1 M NaOH for hydrolysis, or alternatively add 4 µl H_2O as negative control, and incubate at 65° C in a water bath for 1 h. Then move from the water bath to ice.
3. Neutralize with 2 µl of 1.2 M HCl, 4 µl of 1 M Tris-HCl, and 4 µl of H_2O, or alternatively 6 µl of H_2O and 4 µl of 1 M Tris-HCl for negative control. Keep on ice till transformation.

3.3. MATERIALS

3.3.1. Transformation reagents and media (modified from Storici and Resnick, 2006 [116])

1. YPD (Yeast Peptone Dextrose): For 1 L, 10 g yeast extract, 20 g soy peptone, 20 g dextrose. Add 15 g agar to make YPD solid media. Autoclave before use. Store at room temperature.
2. Solution 1: 0.1 M of lithium acetate. Prepare immediately before transformation. Solution 1 is a working solution, thus it is prepared directly from the powder. No stock solution is made. Keep at room temperature. LiAc increases the yeast cell wall permeability to DNA.
3. Solution 2: 0.1 M of lithium acetate and 50 % of polyethylene glycol 4000. Also solution 2 is a working solution and it is made directly from the powder. No stock solution is prepared. Keep at room temperature. PEG deposits oligos onto yeast cells.
4. RNA-containing oligos (Thermo Scientific Dharmacon), 50-80-mers, desalted, deprotected and non-purified. Resuspend to 250 pmoles/µl. Store at -80°C.
5. DNA-only oligos, 50-80-mers, desalted and non-purified. Resuspend to 50 pmoles/µl. Store at -20°C.
6. SC-Trp (Synthetic complete media lacking tryptophan) solid media.
7. 0.5 cm diameter glass beads, sterilized by autoclaving.
8. RNase-off: RNase decontamination solution.
9. DNase/RNase-free, sterile centrifuge tubes.
10. DNase/RNase-free, sterile conical tubes.

11. DNase/RNase-free, sterile aerosol pipette tips with ZAP: 1-200 ml, 100-1000 ml.

3.3.2. Colony PCR materials.

1. DNA primers, desalted and non-purified. Dissolve in sterile water to 50 pmoles/µl. Store at -20°C.
2. Taq DNA polymerase, buffer, dNTPs.
3. PCR tubes.

3.3.3. PCR purification.

1. PCR purification kit.

3.3.4. Gel Electrophoresis.

1. Agarose.
2. 1 x TBE running buffer (45 mM Tris-borate and 1 mM ethylenediamine tetraacetate) diluted from 10x TBE.
3. Prestained molecular weight marker.
4. DNA loading dye.

3.3.5. Restriction digestion.

1. Restriction enzymes, 10x buffers, BSA.

3.3.6. Alkali treatment for the RNA-containing oligo.

1. 1 M of NaOH solution.

2. 1.2 M of HCl solution.
3. 1 M of Tris-HCl, pH 7.4 solution.

3.4. REPRESENTATIVE RESULTS

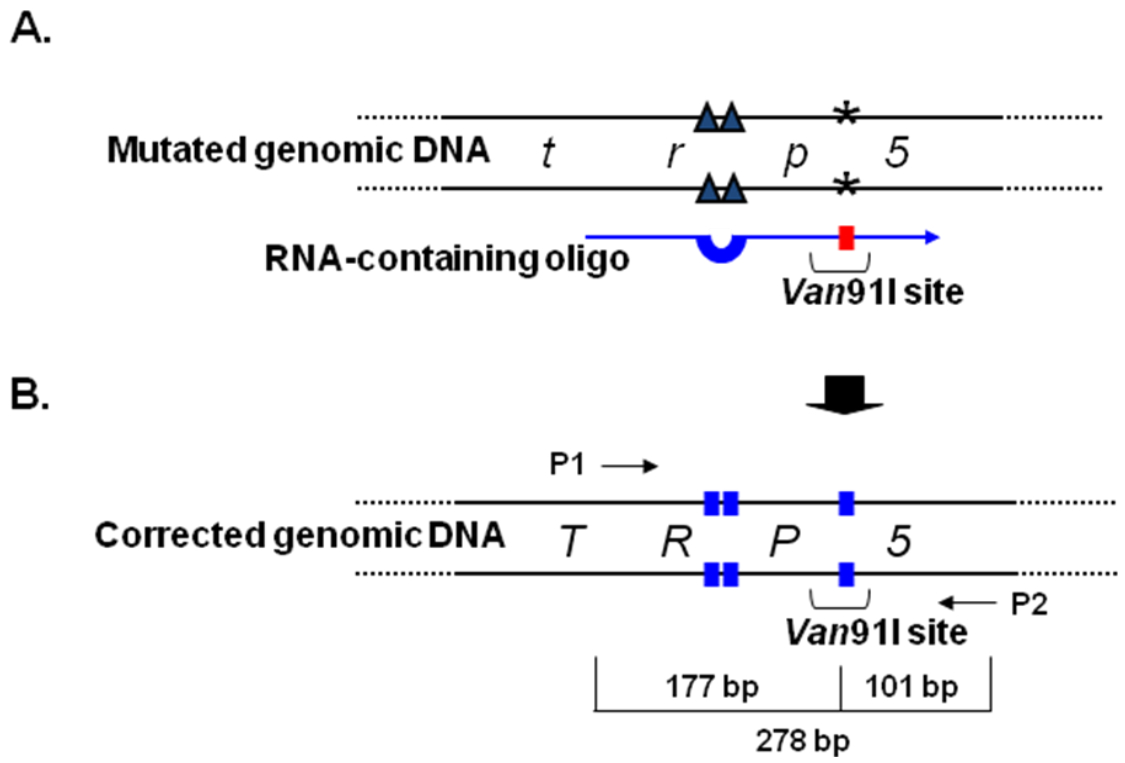


Figure 3.1 Schematic diagram of the defective *trp5* gene and of the *TRP5* allele corrected by the RNA-containing oligo. A) The *trp5* mutant gene contains a 2-base deletion (black triangles) and 1-base nonsense mutation (asterisk). The single-strand RNA-containing oligo with 2-base insertion (blue loop) and 1-RNA base substitution (red rectangle) generating a *Van*91I restriction enzyme site (shown by the bracket) is transformed into yeast cells to correct the genetic defects of the *trp5* gene. B) After the *TRP5* gene is repaired by the RNA-containing oligo (corrected bases are indicated as blue rectangles), the *Van*91I site is generated into the *TRP5* gene. A 278 bp fragment including only one *Van*91I restriction site in *TRP5* gene is PCR amplified by a pair of primers (P1 and P2). The 177 bp and 101 bp fragments generated after digestion of the P1 and P2 PCR product by *Van*91I are also shown.

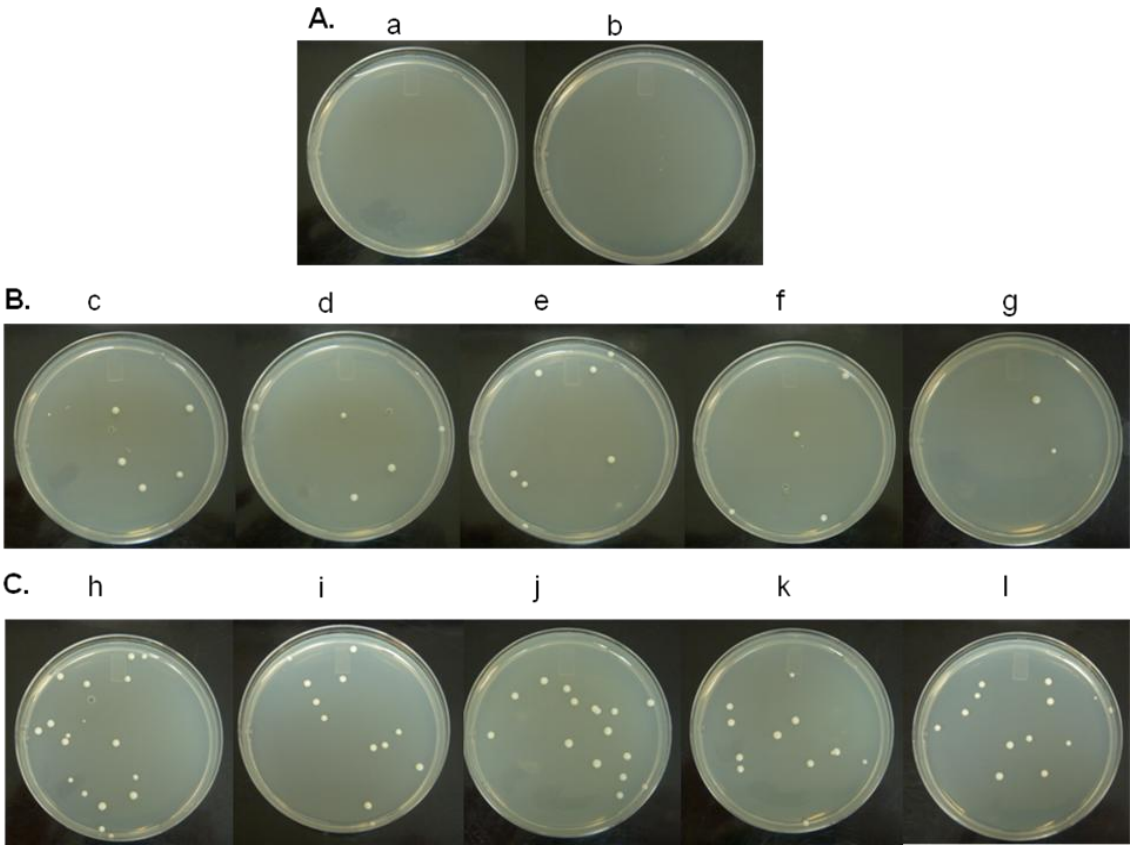


Figure 3.2 Transformation result with the RNA-containing oligo. A) Yeast cells transformed with no oligo do not form any colony on the SC-Trp medium, see plates a-b. B) Plates c-g show yeast colonies growing on SC-Trp after cells are transformed with 1 nmole of the RNA-containing oligo. C) Plates h-l show yeast colonies growing on SC-Trp after cells are transformed with 1 nmole of the corresponding DNA-only control oligo.

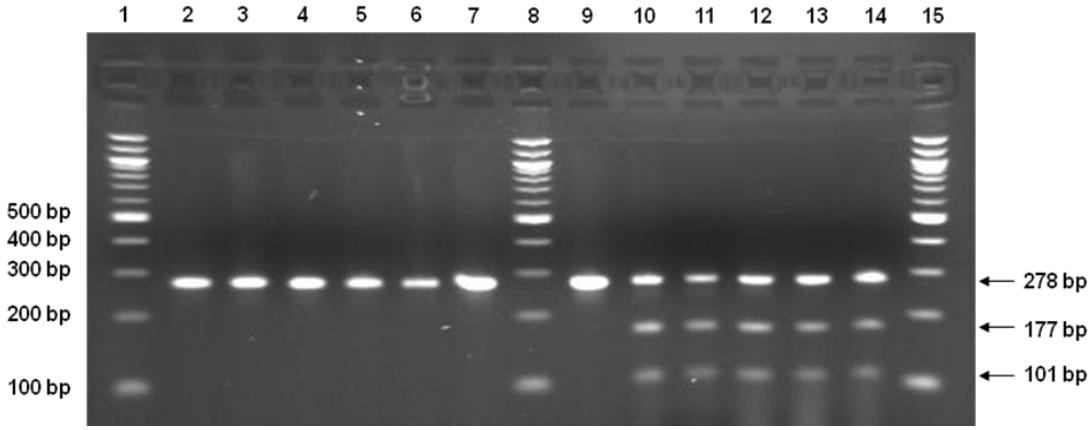


Figure 3.3 Detection of genetic information transfer from the RNA-containing oligo to yeast chromosomal DNA by restriction digestion of the PCR product amplifying the targeted genomic region. 2% agarose gel electrophoresis of PCR samples amplified by P1 and P2 and digested with *Van*91I restriction enzyme. Lanes 1, 8 and 15, DNA ladder with sizes of 100, 200, 300, 400 and 500 bp indicated on the left; lane 2, PCR product of *trp5* locus amplified from the genomic DNA of the *trp5* mutant strain; lane 3, PCR product amplified from genomic DNA derived from a Trp^+ colony targeted by the DNA-only oligo; lane 4-7, PCR products amplified from genomic DNA derived from Trp^+ colonies targeted by the RNA-containing oligo; lane 9 to 14, *Van*91I restriction digestion of the PCR products from lanes 2 to 7. The presence of the uncut PCR product bands in lanes 10 to 14 can be explained by partial digestion by *Van*91I at the *TRP5* locus (CCACATTCTGG). Considering that the cutting site for *Van*91I (CCANNNN'NTGG) can have multiple sequences, the site generated in *TRP5* may not be the most optimal target for the enzyme. In fact, following DNA sequencing of all the above PCR products we detect no additional changes beyond those carried by the oligos (see **Figure 3.4**). The 278 bp PCR band amplified by P1 and P2 primers and the digestion product bands by *Van*91I of 177 bp and 101 bp are shown by the arrows on the right.

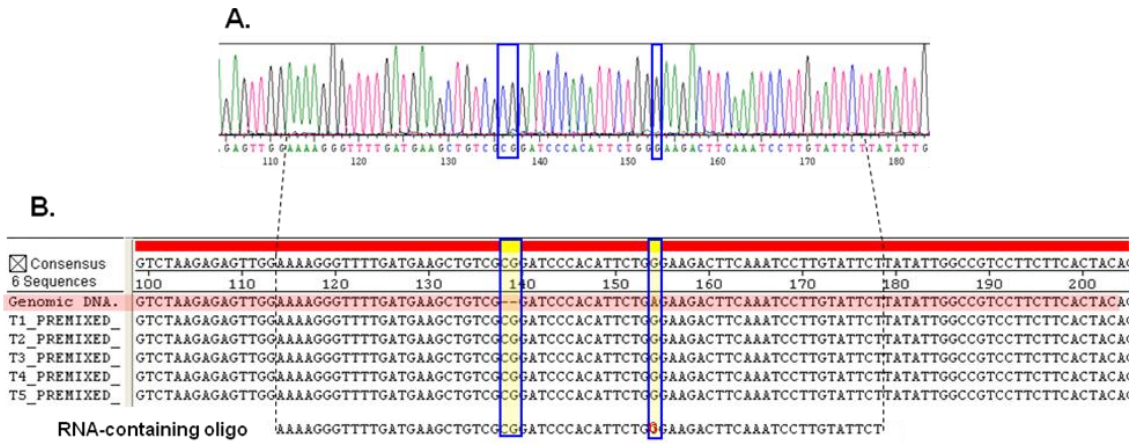


Figure 3.4 DNA sequencing results showing gene correction by the RNA-containing oligo. A) DNA electropherogram of the genomic region targeted by the RNA-containing oligo. The G in the DNA sequence (blue boxed) derives from the rG on the RNA-containing oligo. Also boxed is the insertion of the CG bases. Sequencing results from all other PCR products are also as clear as this one, with fluorescent signals well above the background. B) Sequences of the *TRP5* region of the Trp^+ transformants targeted by the DNA-only oligo (T1) and the RNA-containing oligo (T2-T5) match in the consensus sequence on the top and are compared with that of the *trp5* mutant cells before targeting by the oligos (Genomic DNA, red shaded). The repair RNA-containing oligo on the bottom has two-base DNA insertion and one-base (G) RNA substitution marked in red. The regions boxed in blue with yellow shade show that the RNA-containing oligo as well as the DNA-only oligo precisely corrected the deletion mutation and nonsense mutation in all tested samples. The dashed lines mark the position of the RNA-containing oligo sequence.

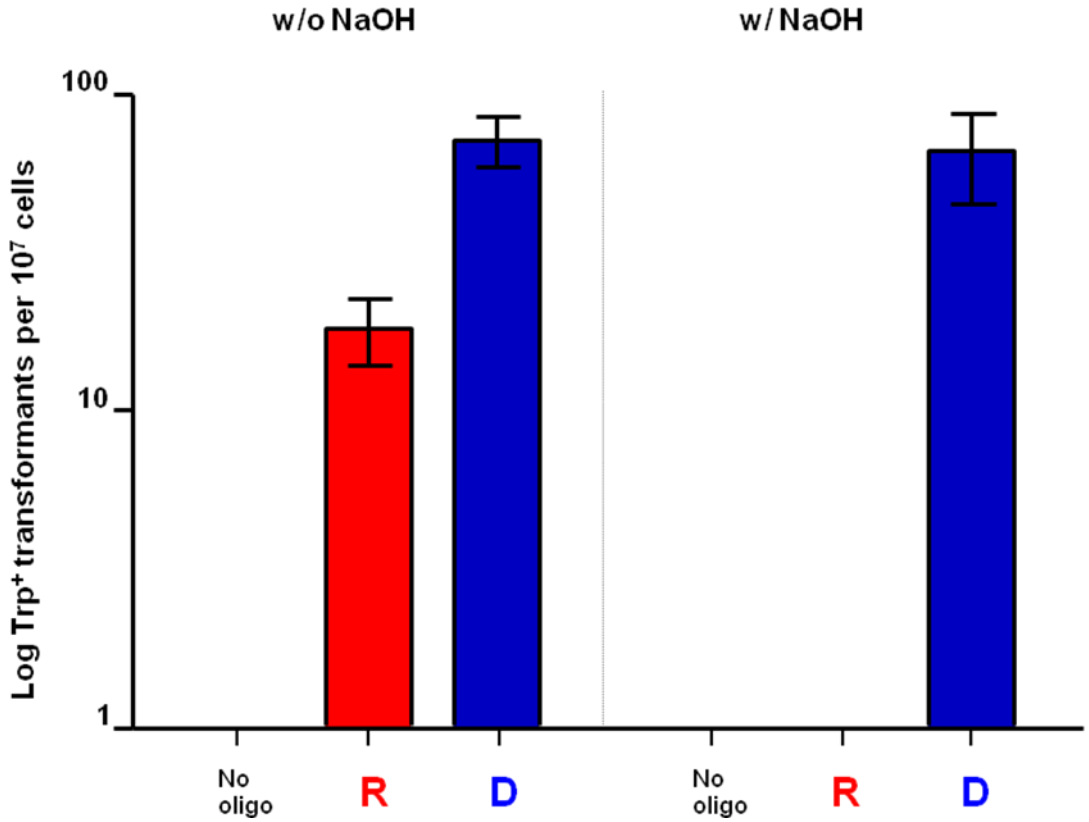


Figure 3.5 Alkali treatment prevents gene correction by the RNA-containing oligo. The transformation frequency by the RNA-containing oligo (R) is shown in the red bar and that by the DNA-only oligo (D) is shown in the blue bars. The error bars represent the standard deviations of the data for 3 independent transformations for each oligo. The DNA-only oligo displays similar transformation frequency without and with NaOH treatment. Differently, transformation frequency by the RNA-containing oligo drops to <0.1 following treatment with NaOH. Therefore, the preparation with the RNA-containing oligo is not contaminated with DNA-only oligo, thus, the observed transformation frequency is specific to the RNA-containing oligo.

3.5. DISCUSSION

The fact that RNA can transfer genetic information directly to genomic DNA in cells was discovered exploiting the use of synthetic RNA-containing oligos (Dharmacon synthesized oligos) [77]. It was the proof of principle that cells can use RNA-containing molecules or RNA-only sequences as templates for DNA synthesis. The use of RNA-containing oligos not only led to the demonstration that genetic information can be transferred directly from RNA to genomic DNA without the need of a reverse transcribed DNA copy intermediate, but also to the proof that RNA can be used as homologous template in the repair of a DNA damage [77, 91]. The oligos can be made of RNA-only or contain just a single ribonucleotide embedded in a DNA sequence, as in the example we presented here. The RNA-containing oligo has one embedded ribonucleotide designed to correct the nonsense mutation in the genomic defective *trp5* allele. The transfer of genetic information from the ribonucleotide to the genomic DNA is revealed by a change in phenotype (cell growing on the medium lacking tryptophan) of the targeted cells. The frequency of gene correction obtained with the RNA-containing oligo is measured and compared with that of a control DNA-only oligo. By performing this gene correction assay in different cell backgrounds, where we mutate various yeast genes, we can detect which factor/s specifically affect targeting by the RNA-containing oligos. Thus, we can reveal mechanisms how cells regulate the stability of RNA/DNA hybrids. By simply designing different variants of the RNA-containing oligos we can determine the likelihood for a particular RNA tract to serve as template in DNA modification and we can determine the stability of specific RNA sequences embedded into DNA. Moreover,

we can identify what are the preferred *in vivo* substrates for factors interacting with RNA/DNA hybrids.

While several *in vitro* studies, mainly utilizing short RNA-containing oligos, have been conducted to characterize the function of factors that can recognize RNA in a hybrid with DNA, such as the RNase H enzymes [20], the *in vivo* functions of RNases H as well as the identity of other proteins that could affect RNA/DNA hybrid stability remain mostly unknown. The possibility to utilize RNA-containing oligos of a significant length (50 to 80-mers) and optimal quality (such as Thermo Scientific Dharmacon RNA-containing oligos) has open the way to examining a wide range of molecular processes directly *in vivo* in the cells of interest. As shown in this video, transformation using RNA-containing oligos requires essentially only a few additional steps compared to transformation using DNA molecules, to prevent degradation by RNases. Thus, transformation using RNA-containing oligos is not limited to the yeast system, but can be applied to any organism or cell type where transformation by DNA oligos is proficient.

In conclusion, targeting cells by RNA-containing oligos provides the opportunity to generate RNA/DNA hybrids and ribonucleotides embedded in DNA *in vivo* in cells. The stability, function and consequences of these *in vivo* generated RNA/DNA hybrids can be analyzed and characterized, potentially uncovering unknown mechanisms of DNA repair and revealing novel strategies for gene targeting.

3.6. ACKNOWLEDGEMENTS

This work was supported by the Georgia Cancer Coalition grant-R9028.

.

CHAPTER 4

Mispaired rNMPs in DNA are mutagenic and are targets of mismatch repair and RNases H

The paper of the study in Chapter 4 is in press in Nat. Struct. & Mol. Biol., 2011.

Shen, Y.[1], Koh, K.D.[1], Weiss, B.[2] and Storici, F.[1]

[1]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230;

[2]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322.

*The RNase HII cleavage experiments and biochemical data analysis were done by Koh, K.D.

4.1. ABSTRACT

Numerous studies have shown that ribonucleoside monophosphates (rNMPs) are abundant among all non-standard nucleotides occurring in genomic DNA. Therefore, it is important to understand to what extent rNMPs may alter genome integrity and what factors affect their stability. We developed oligonucleotide-driven gene correction assays in the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae* to show that mispaired rNMPs embedded into genomic DNA, if not removed, serve as templates for chromosomal DNA synthesis and produce a genetic change. We discovered that isolated mispaired rNMPs in chromosomal DNA are removed by the mismatch repair system in competition with RNase H type 2. However, a mismatch within an RNA/DNA heteroduplex region requires RNase H type 1 for efficient removal. In the absence of mismatch repair and RNase H functions, ribonucleotide-driven gene modification increased 47-fold in yeast and 77,000-fold in *E. coli*.

4.2. INTRODUCTION

Modifications of nucleotides in DNA pose a threat to the genomic integrity of cells, often resulting in cell death or mutation. Numerous studies suggest that ribonucleoside monophosphates (rNMPs) are abundant among all non-standard nucleotides occurring in genomic DNA. Many DNA polymerases have been shown *in vitro* to incorporate rNMPs into DNA [117-122], including all yeast replicative polymerases [123]. Importantly, the combination of ribonucleotide misincorporation and mispairing *in vivo* can occur with sufficient frequency to affect mutation rates in yeast [111, 124]. The finding that rNMP incorporation during replication in yeast is coupled with genome instability in an RNase H2 defective background [111] raises the question whether mismatches generated by mispaired rNMPs or processing of rNMPs in DNA are subject to mismatch repair (MMR).

The concentration of ribonucleoside triphosphates in cells is generally higher than that of deoxyribonucleoside triphosphates, increasing the probability of paired and mispaired rNMP incorporation by DNA polymerases during DNA replication and repair ([123] and references therein). Regions of highly transcribed DNA were found to contain a high concentration of dUMP in yeast [7], and it would be interesting to see if these regions may also be hot spots for rNMP incorporation. In addition, DNA primases have very low fidelity, resulting in the incorporation of mispaired rNMPs, and could potentially leave one or several rNMPs embedded in chromosomal DNA by including deoxyribonucleotides in addition to ribonucleotides in the primer sequence [8, 9]. Finally, oxidative damage of DNA can convert a deoxynucleotide in DNA to an rNMP [10].

Ribonucleotides in DNA can distort the double helix [11], resulting in genomic instability, defective replication [112] or transcription and mutagenesis. Topoisomerase I can initiate a process of rNMP removal when its site of cleavage occurs 3' to an rNMP embedded in DNA [125]. However, the enzymes that specifically cleave RNA in RNA/DNA hybrids are ribonucleases H (RNase H type 1 (I or 1) and type 2 (II or 2)) [20, 126]. *In vitro*, RNases HI/1 and RNases HII/2 have very distinct cleavage patterns. RNases HI/1 require a substrate with an RNA stretch containing at least four ribonucleotides in a DNA duplex to allow cleavage, while RNases HII/2 can cleave a single ribonucleotide embedded in DNA even when mispaired [20, 23]. Yet, up to now a detailed analysis of the *in vivo* substrate specificity of these enzymes is missing [20].

4.3. MATERIALS AND METHODS

4.3.1. Bacterial strains. All *E. coli* strains used here (**Table A.1a)** have a thermoinducible λ *red* recombination system, which permits the efficient incorporation of linear DNA into the chromosome [100]. Because both RecB and RecC are required for the viability of an *rnhA* mutant [127], we deleted the prophage *N-gam* segment [100], the product of which is an inhibitor of the RecBCD enzyme [128]. The partially deleted prophage retained functional *exo* and *bet* genes of the λ *red* region, of which only the *bet* gene is necessary for oligo transformation [100]. Transductions were performed with bacteriophage P1 *dam rev6* [129]. We used both deletion and nonsense mutations for *rnhA* and *rnhB*. Complete deletions of the coding regions are ideal because they eliminate the possibility that defective products might still be present, although they also remove the promoters of neighboring *dnaQ* and *dnaE* genes, which then are transcribed, we assume, from alternative promoters. Therefore, we also constructed truncated forms of RNase HI and HII (verified by sequencing) by introducing two consecutive stop codons (Ochre and Opal) upstream of *dnaQ* and *dnaE* promoters and after codon 48 in *rnhA* and after codon 70 in *rnhB* in strain BW1988 and BW2037 exploiting oligo-directed transformation (**Table A.1a**). We then compared the transformation frequencies by the *LacZ.R6*$_{I2}$, *LacZ.R1*$_{S1}$ or *LacZ.D* oligo in the *rnh* deletion and nonsense mutants and no differences were revealed (**Table A.6a, b**). We concluded that both the deletion and nonsense *rnh* mutant alleles behave similarly, at least with respect to gene correction by RNA-containing oligos. The *mutS*::*gm* mutation was replaced in some strains by the *mutS*$^{+}$ allele, which was cotransduced with a nearby Tn*10* in a gene for sorbitol utilization

(*srlD*). The *mutS*$^{+}$ and *mutS* transductants were scored by the relative frequency of spontaneous mutation to streptomycin resistance. Bacterial media and growth conditions were as described before [130].

4.3.2. Yeast strains. The yeast haploid strains used in this work derives from FRO-694, which contains the GSKU cassette [65] and the I-SceI cutting site in *TRP5,* inserted between nucleotide (nt) C1001 and nt C1002 (**Table A.1b**). Three different oligos with homology to the *TPR5* gene containing a 2-base deletion (ΔC-C1001-1002) plus a nonsense mutation (nt G1017 to A), only the 2-base deletion mutation, or only the nonsense mutation were used to transform the FRO-694 to pop out the GSKU cassette in *trp5* and generate the three different *trp5* alleles in three different strains, YS-301/303, YS-316/318 and YS-320/322, respectively (**Table A.1b**), following the "*delitto perfetto*" method as described [65]. The oligos used for the construction of the different *trp5* alleles are shown here with the introduced mutations in bold:

80DeltaCC+STOP_TRP5.e:

5'-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCG__GATCCCACATTCTG

AGAAGACTTCAAATCCTTGTATTCTTATATT (2-base deletion & nonsense mutation)

80DeltaCC_TRP5.e:

5'-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCG__GATCCCACATTCTG

GGAAGACTTCAAATCCTTGTATTCTTATATT (2-base deletion mutation)

82STOP_TRP5.e:

5'-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCGCCGATCCCACATTCTG

AGAAGACTTCAAATCCTTGTATTCTTATATT (nonsense mutation).

Null alleles for *RNH201*, *MSH2* or both were created by replacing the open reading frame of each of these genes by either the *kanMX4* or the *hygMX4* module.

4.3.3. Transformation using oligos. For experiments in *E. coli*, electrocompetent cells were prepared as described [99] and used immediately. The final wash and resuspension were in RNase-free water. RNA-containing oligos retained ≥90% of their activity after incubation with fresh cells for 3 min. Electroporation was performed as described [99], immediately after adding 100 ng of oligo in 50 μl of a cell suspension. For transformation experiments to mutate the $rpsL^+$ gene, 10 μl of cultures after recovery in the SOC media were transferred to 1 ml of LB media and incubated overnight to reduce the proportion of the original streptomycin-sensitive ribosomes. Cells were then diluted in 10 mM $MgSO_4$ and plated in an overlay of 4 ml of soft agar on LB agar with streptomycin to select for transformants and on LB agar to measure viability. Survival after transformation was 24%-30% for BW1988, BW2037 and all the derivative strains used. All strains with *rnhA* mutations form smaller colonies on LB agar media than strains with wild-type *rnhA*. For experiments in yeast, transformation with RNA-containing or DNA-only oligos (1 nmole) were done as described [77]. Cells from each oligo transformation were plated to selective Trp^- media and were diluted and plated on the rich YPD media. Survival after yeast transformation was 29% for WT, 26% for *msh2*, 24% for *rnh201* and 29% for *rnh201 msh2* cells, respectively. The relative transformation frequencies were calculated by dividing the number of transformants per 10^7 viable cells obtained with an RNA-containing oligo by the median of the number of transformants per 10^7 viable cells

obtained with the corresponding DNA-only oligo in the same experiment. The results are each expressed as a median and 95% confidence limits (in parentheses), or alternatively the range when number of repeated experiments was <6. Random clones derived from gene correction by *LacZ.R1*$_{S1}$ in the *rnhB mutS* background, *LacZ.R5*$_{S1}$ in the *rnhA rnhB* background, *RpsL.R1*$_{S1}$ in the *rnhB mutS* background, or *TRP5.R2_R1*$_{I2\_S1}$ in the *rnh201 msh2* background were sequenced at the region targeted by the oligos, and all (24/24, 19/19, 13/13 and 14/14, respectively) had the expected corrected *lacZ*$^{+}$, *rpsL*^{r}, or *TRP5* sequence, with no additional changes.

4.3.4. Alkali treatment of oligos. Two μl of a 10 μg/μl solution for RNA-containing or DNA only oligos were treated before the transformation experiments in *E. coli* with 2 μl of 1 M NaOH at 65° C for 1 h, and then neutralized with 0.2 μl of 1 M HCl and 0.8 μl of 1 M Tris-HCl pH 7.4, and H_2O was added to 100 μl. Alternatively, 2 μl of the same RNA-containing or DNA-only oligo solutions were treated with 2 μl of H_2O at 65° C for 1 h, and then neutralized with 2 μl of H_2O and 0.8 μl of 1 M Tris-HCl pH 7.4, and H_2O was added to 100 μl. Alkali treatment of oligos for experiments in yeast was performed as described [77].

4.3.5. Standard genetic and molecular biology techniques. Standard genetics and molecular biology analyses were done as described [65, 77, 99]. Samples for sequencing were submitted to Eurofins MWG Operon, (Huntsville, AL).

4.3.6. RNase HII cleavage assay. RNA-containing oligos and their corresponding DNA-only oligos (**Table A.2**) were 5'-end-labeled using [γ-32P]ATP (PerkinElmer, Waltham, MA) and T4 Polynucleotide Kinase (NEBioLabs, Ipswich, MA). Double-stranded substrates (**Figure 4.2**) were prepared by annealing appropriate complementary oligos in 1X ThermoPol Reaction Buffer (NEBioLabs), which is the RNase HII reaction buffer, at pH 8.8. RNase HII reactions were performed by incubating 40 fmol of each substrate with 1 unit of RNase HII (NEBioLabs) in the same buffer for 2 hours at 37°C, or with 0.75 units of RNase HII for 30 minutes at 37°C. 20/100 Oligo Length Standard (IDT, Coralville, IW) was used as a marker. Gel electrophoresis was run with a denaturing 15% polyacrylamide (19:1) / Urea (8M) gel at 50W with 1X TBE buffer at 50°C. Following electrophoresis, gels were exposed to a phosphor screen overnight. Images were taken with Typhoon Trio^{+} (GE Healthcare, Piscataway, NJ). Band intensities were quantified by Multi Gauge V3.0 (FUJIFILM, Tokyo, Japan).

4.4. RESULTS

4.4.1. Impact of RNase HI and RNase HII on gene correction by rNMP-containing oligos in mismatch repair defective *E. coli*

We initially set up a gene correction assay for different RNA-containing oligonucleotides (oligos) in the *E. coli* strain BW1988, which is an MMR-deficient (*mutS* null) strain expressing the λ phage recombination system for highly efficient oligo targeting [99, 105] (**Table A.1a**). In this system, lagging-strand oligos anneal to the complementary gapped single-strand region at the replication fork, are incorporated into the genome similarly to Okazaki fragments and serve as templates for gene correction in the next round of DNA replication [105]. A lagging-strand oligo containing a tract of 6 rNMPs was also much more efficient at gene correction than the corresponding complementary oligo [67]. The *E. coli* BW1988 strain had a *lacZ* marker gene altered by a two-base frameshift deletion (ΔGG1370-1) and a missense mutation at nucleotide 1384, changing codon GAG (Lac^+) into AAG (Lac^-) [99] (**Table A.1a**). To correct these *lacZ* mutations, BW1988 was transformed with lagging–strand oligos having the sequence of the wild-type *lacZ* gene and containing 6 or 2 rNMPs opposite to the deletion (*LacZ.R6*$_{I2}$ or *LacZ.R2.47*$_{I2}$), 1 rNMP or 5 rNMPs opposite to the missense mutation (*LacZ.R1*$_{S1}$ or *LacZ.R5*$_{S1}$), or no rNMPs (*LacZ.D* or *LacZ.D.47*) (**Figure 4.1a** and **Table A.2**). Transformation results are shown in **Table 4.1**. In the *mutS* background, differently from the *LacZ.R6*$_{I2}$ and the *LacZ.R2.47*$_{I2}$, the *LacZ.R1*$_{S1}$ and the *LacZ.R5*$_{S1}$ oligos produced a surprisingly low frequency of Lac^+ transformants relative to the control DNA-only oligos. The experiment was repeated with mutants for RNase HI (*rnhA*) and RNase HII (*rnhB*) in the *mutS* background to see if the enzymes would affect oligo transformation by

attacking the embedded rNMPs. As presented in **Table 4.1**, the *rnhA* mutation showed a strong effect only on the transformation frequency by *LacZ.R5$_{S1}$* oligo (85-fold ($6*10^{-3}/7*10^{-5}$) increase). Differently, the *rnhB* mutation enhanced the relative transformation frequency of the *LacZ.R1$_{S1}$* oligo 6,800-fold ($3.43/5*10^{-4}$) and that of the *LacZ.R6$_{I2}$* oligo 8-fold (0.16/0.02). It had no effect on the *LacZ.R2.47$_{I2}$* oligo, and it enhanced the frequency of the *LacZ.R5$_{S1}$* oligo only in combination with the *rnhA* mutation. Clearly, our data show that *in vivo* RNase HII very efficiently targets a single mispaired rNMP and cannot target the isolated 2-rNMP loop, whereas RNase HI is preferred to RNase HII for targeting a tract of 5 rNMPs with a mispair. The stronger resistance of *LacZ.R6$_{I2}$* and in particular of *LacZ.R2.47$_{I2}$* to cleavage by RNase HI and HII is probably due to the secondary structure of the insertion loop present in these oligo sequences. From *in vitro* studies it is known that RNases HI/1 can cleave DNA substrates containing a stretch of at least four ribonucleotides, whereas RNases HII/2 can cleave a single ribonucleotide embedded in DNA, even when mispaired, or longer embedded rNMP tracts[20, 23]. It was then surprising to find that the *rnhB* mutation did not affect the gene correction frequency by the *LacZ.R5$_{S1}$* and the *LacZ.R2.47$_{I2}$* oligos and it only modestly affected that by the *LacZ.R6$_{I2}$* oligo compared to that by the *LacZ.R1$_{S1}$* oligo (**Table 4.1**). The data also demonstrate that all the different rNMPs embedded in double-stranded *E. coli* chromosomal DNA are used as templates for DNA synthesis.

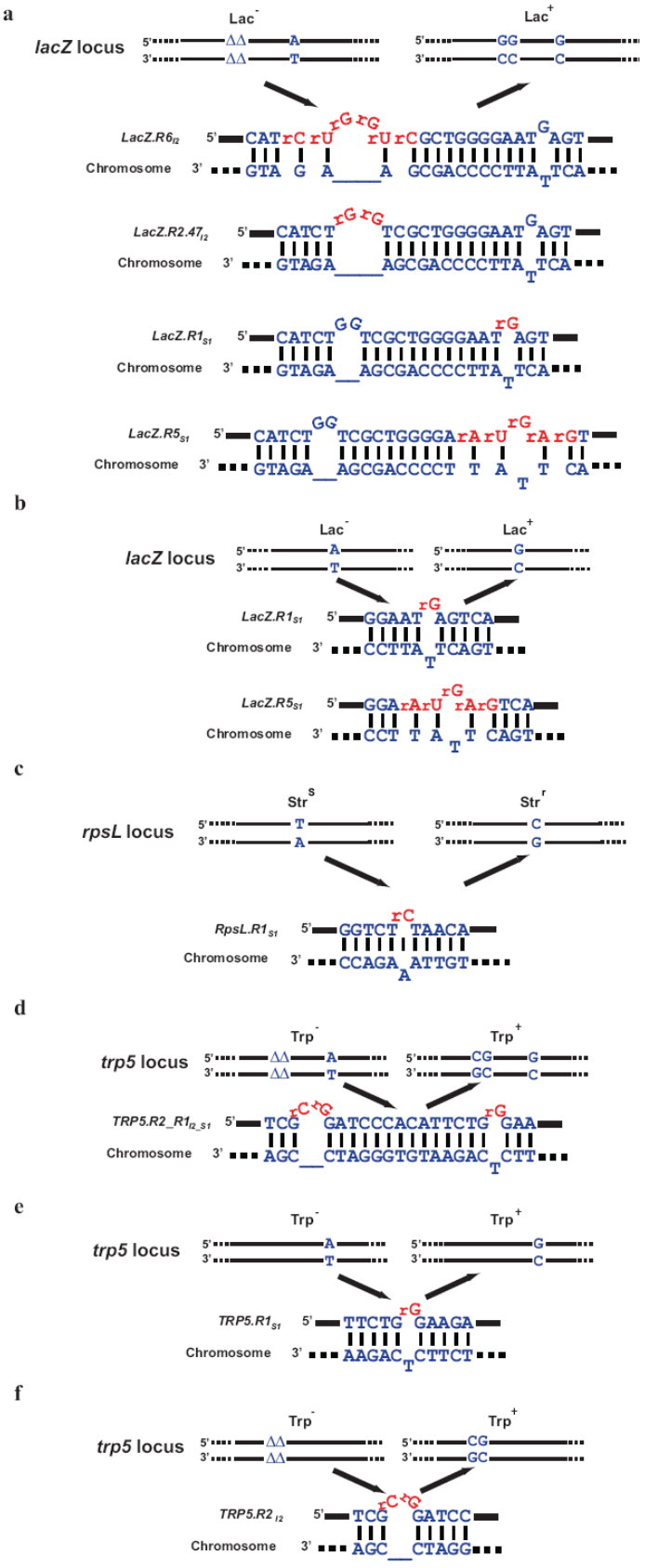

a
lacZ locus
Lac-
Lac+
LacZ.R6I2
LacZ.R2.47I2
LacZ.R1S1
LacZ.R5S1
Chromosome
b
lacZ locus
Lac-
Lac+
LacZ.R1S1
LacZ.R5S1
Chromosome
c
rpsL locus
Str s
Str r
RpsL.R1S1
Chromosome
d
trp5 locus
Trp-
Trp+
TRP5.R2_R1I2_S1
Chromosome
e
trp5 locus
Trp-
Trp+
TRP5.R1S1
Chromosome
f
trp5 locus
Trp-
Trp+
TRP5.R2I2
Chromosome

Figure 4.1 Diagrams and sequences of the *E. coli lacZ* and *rpsL* loci and the yeast *trp5* locus targeted by the RNA-containing oligos. In the name of the RNA-containing oligos, substitutions are indicated by a subscript capital "S" and insertions by a subscript capital "I". The letters "S" and "I" are followed by a subscript number indicating the number of bases that are substituted or inserted, respectively. (**a**) The *lacZ* locus containing a two-base deletion and a substitution mutation targeted by the *LacZ.R6*$_{I2}$, *LacZ.R2.47*$_{I2}$, *LacZ.R1*$_{S1}$, or *LacZ.R5*$_{S1}$ oligo. (**b**) The *lacZ* locus containing a substitution mutation targeted by the *LacZ.R1*$_{S1}$, or *LacZ.R5*$_{S1}$ oligo. (**c**) The *rpsL* locus targeted by the *RpsL.R1*$_{S1}$ oligo. (**d**) The *trp5* locus containing a two-base deletion and a substitution mutations targeted by the *TRP5.R2_R1*$_{I2\_S1}$ oligo, (**e**) containing just a substitution mutation targeted by the *TRP5.R1*$_{S1}$ oligo, or (**f**) containing only a two-base deletion mutation targeted by the *TRP5.R2*$_{I2}$ oligo.

Table 4.1 The effect of RNase HI (*rnhA*) and RNase HII (*rnhB*) mutations on transformation by rNMP-containing oligos in *E. coli mutS* cells

Genotype	*LacZ.R6$_{I2}$*		*LacZ.R2.47$_{I2}$*		*LacZ.R1$_{S1}$*		*LacZ.R5$_{S1}$*	
mutS	0.02	(0.01-0.03)	0.18	(0.13-0.22)	$5*10^{-4}$	($4*10^{-4}$-$6*10^{-4}$)	$7*10^{-5}$	(0-$9.7*10^{-5}$)
mutS rnhA	0.04	(0.02-0.05)	ND		$9*10^{-4}$	($7*10^{-4}$-$1.4*10^{-3}$)	$6*10^{-3}$	($5.5*10^{-3}$-$7.3*10^{-3}$)
mutS rnhB	0.16	(0.13-0.18)	0.16	(0.13-0.20)	3.43	(1.74-3.87)	$9*10^{-5}$	($4.3*10^{-5}$-$1.8*10^{-4}$)
mutS rnhA rnhB	0.38	(0.35-0.46)	ND		2.40	(1.94-2.81)	0.08	(0.06-0.09)

The values are relative frequencies of Lac^+ transformants (see Methods). The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney test ($P < 0.05$) (**Table A.3a**). The median transformation frequencies per 10^7 cells and 95% confidence limits (in parentheses) for each of the strains with the *LacZ.D* and the *LacZ.D.47* oligos were as follows: *mutS*, 7,290 (4,120-10,800); *mutS rnhA*, 7,780 (5,090-11,300); *mutS rnhB* 4,840 (2,880-6,950); *mutS rnhA rnhB*, 12,908 (7,750-18,000), and *mutS*, 2,430 (1,770-3,760); *mutS rnhB* 2,720 (2,250-3,630), respectively. In the absence of a transforming oligo, the frequency of Lac^+ colonies per 10^7 viable cells were <0.1. The numbers of repeats for each of the *mutS*, *mutS rnhA*, *mutS rnhB*, and *mutS rnhA rnhB* strains transformed with these oligos were as follows: *LacZ.R6$_{I2}$:* 7, 7, 6, 7; *LacZ.R2.47$_{I2}$*: 4, ND, 4, ND; *LacZ.R1$_{S1}$*:7, 7, 7, 6; *LacZ.R5$_{S1}$:* 6, 6, 6, 6; *LacZ.D*: 6, 6, 6, 6; *LacZ.D.47*: 4, ND, 4, ND; no oligo: 10, 10, 10, 10. ND, not determined. The strains used were BW1988, BW2028, BW2029, and BW2032.

4.4.2. RNase HII cleavage of heteroduplexes containing paired or mispaired rNMPs

The inability of the *rnhB* mutation to increase gene correction frequencies by the *LacZ.R5*$_{S1}$ and the *LacZ.R2.47*$_{I2}$ oligos and its capacity to increase gene correction by the *LacZ.R6*$_{I2}$ oligo to a much lesser extent than by the *LacZ.R1*$_{S1}$ oligo may be a reflection of the specificity of RNase HII. To test this hypothesis, we analyzed the ability of the purified enzyme to cleave heteroduplexes containing the same sequences and RNA/DNA mismatches that we examined in the transformation experiments (**Table A.2** and **Figure 4.2a**). RNase HII cleavage of these different substrates was determined by polyacrylamide gel electrophoresis (PAGE) following two-hour incubation at 37 ˚C (see Methods section). Under these experimental conditions, substrates with a single rNMP mispair, whether alone or in a tract of 5 rNMPs (substrates S2 and S6, respectively), were cleaved to the same extent as the rNMP-containing substrates with no mismatches (S1 and S5, respectively) (**Figure 4.2a**,**b**). RNase HII cleaved 5' to the last ribonucleotide, or the junction ribonucleotide, in the 5 and 6-ribonucleotide tracts. The faint bands appearing below the major cleavage band for each substrate containing more than 1 rNMP are products of RNA degradation, as they appear also in the absence of RNase HII (**Figure A.1**). The substrate containing a loop of two unpaired rNMPs alone (S11) was not cleaved. The substrate containing the same rNMP loop within a stretch of 6 rNMPs was cleaved to a lesser extent (53 %) than those containing a single mispair (S2 and S6). These data, except those obtained with the S6 substrate, are entirely consistent with our *in vivo* results (**Table 4.1**), in which gene correction by *LacZ.R2.47*$_{I2}$ or the *LacZ.R6*$_{I2}$ oligo is significantly more efficient than correction by the *LacZ.R1*$_{S1}$ oligo in the presence of RNase HII, and deletion of *rnhB* does not increase at all gene correction by the

LacZ.R2.47$_{I2}$ oligo. When we performed a similar experiment in conditions in which RNase HII cleaved the substrate with a single rNMP mispair (S2) at 77 %, the substrate with the same mispair in a tract of 5 rNMPs (S6) was cleaved at 50 %, thus less efficiently than S2 though much more efficiently than S4 (5.4 %) and S11 (<1 %) (**Figure 4.2c**,**d**). Overall, the RNase HII cleavage results do not explain why *in vivo* the *rnhB* mutation did not increase gene correction by the *LacZ.R5*$_{S1}$ oligo while it increased gene correction by the *LacZ.R6*$_{I2}$ (**Table 4.1**). Therefore, in this case, there is no correlation between the preference for the substrate containing a single rNMP mispair in a tract of 5 rNMPs *in vitro* and *in vivo.*

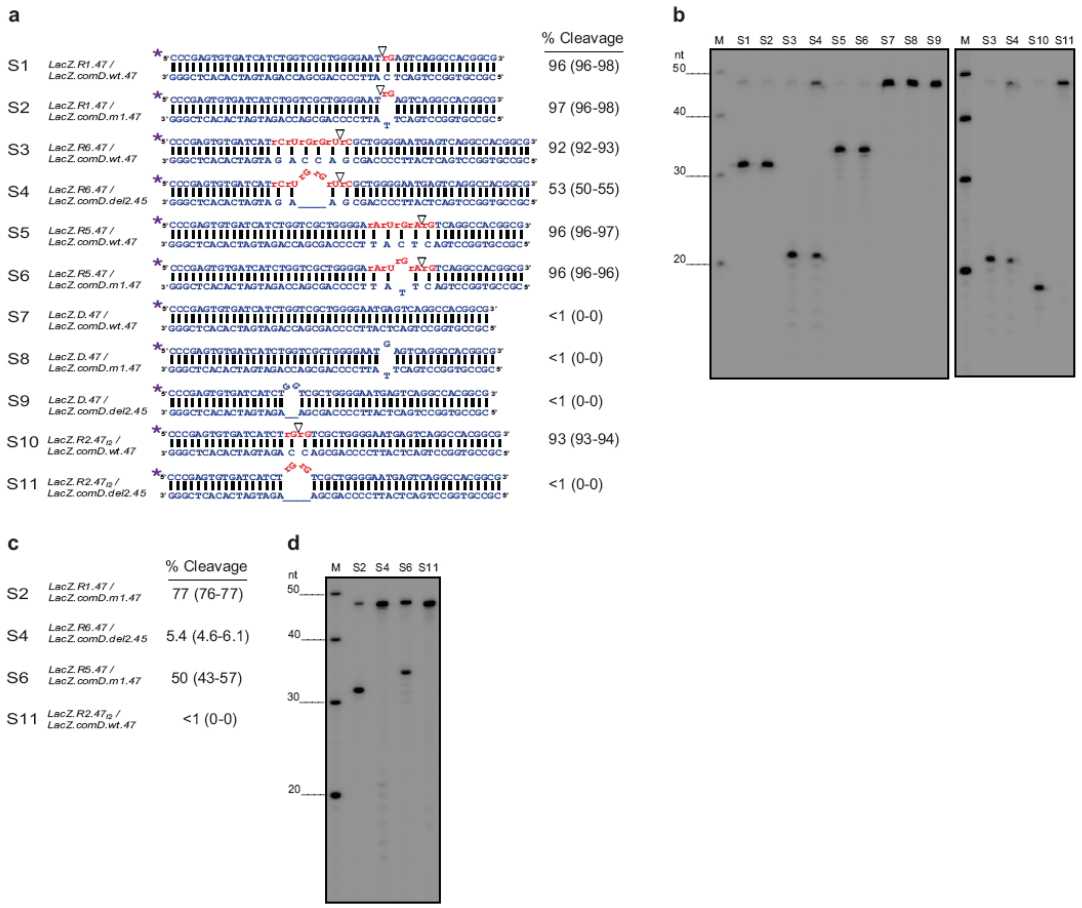


Figure 4.2 RNase HII cleavage specificity. (**a**) Structural presentation of 5'-radiolabeled (^{32}P, indicated by a purple asterisk) substrates (S1-S11) and cleavage percentage for each substrate, expressed as median and range (in parentheses) from 3 independent samples. Inverted triangles indicate the cleavage sites. (**b**) Denaturing polyacrylamide gels showing fragments resulting from cleavage using RNase HII. M: 20/100 nt oligonucleotide marker. S1-S11: substrates used. (**c**) Substrates (S2, S4, S6, S11) used in the experiment shown in (d) and their cleavage percentage, expressed as mean and range (in parentheses) from 2 independent samples. (**d**) Denaturing polyacrylamide gel showing fragments resulting from cleavage using reduced amount of RNase HII and shorter incubation time. M: 20/100 nt oligonucleotide marker. S2, S4, S6, S11: substrates used.

4.4.3. Functional redundancy between mismatch repair and RNase H type 2 functions in the removal of an rG/dT mispair in *E. coli*

We next examined the capacity of the MMR system to recognize RNA/DNA mismatches in *E. coli*. The MMR system is highly conserved from prokaryotes to eukaryotes. It recognizes DNA/DNA mispairs, such as base-base mismatches as well as small insertions and deletions (ins/dels) occurring both during DNA replication [40] and recombination [131]. We modified the bacterial strain BW1947, containing just one missense mutation in *lacZ* (nucleotide 1384 G $\rightarrow$ A) [99], to generate a set of strains having wild-type and mutant alleles of *mutS*, *rnhA* and *rnhB* in all combinations (**Table A.1a**). We then transformed these strains with the *LacZ.R1*$_{S1}$ (**Figure 4.1b**) or the *LacZ.D* control oligo. Transformation results are shown in **Table 4.2**. As expected, *lacZ* correction by the DNA-only oligo in *mutS* cells was several hundred-fold higher than in all *mutS*$^{+}$ cells, in agreement with previous findings [99, 105]. MutS was also able to recognize the mismatched rG/dT mispair, as gene correction frequency by the *LacZ.R1*$_{S1}$ oligo in *mutS* was 5-fold higher than in *mutS*$^{+}$ cells. The relative transformation frequency of *LacZ.R1*$_{S1}$ increased only to 2-fold in the absence of RNase HI both in *mutS*$^{+}$ and in *mutS* cells, in line with the observation that RNase H type 1 cannot cleave at rNMPs in stretches of less than 4 [20, 23]. In *mutS*$^{+}$ cells an *rnhB* mutation enhanced the frequency of transformation by *LacZ.R1*$_{S1}$ only 11-fold (21.8/1.88) (**Table 4.2**), as opposed to the 6,800-fold enhancement seen in a *mutS* mutant (**Table 4.1**). The *rnhB mutS* double mutant showed correction frequency for the *LacZ.R1*$_{S1}$ oligo 17,500-fold (33,000/1.88) higher than the wild type and 1,500-3,400-fold higher than either of the single mutants (**Table 4.2**). The gene correction frequency by *LacZ.R1*$_{S1}$ increased 77,000-fold

(145,000/1.88) in the *rnhA rnhB mutS* triple mutant compared to the frequency in wild-type cells, and became similar to that of the corresponding DNA-only oligo in the *mutS* backgrounds used (**Table 4.2**). Further purification of the *LacZ.R1*$_{S1}$ oligo by PAGE produced no observable changes in gene correction frequencies (not shown).We excluded the possibility of DNA contamination in our *LacZ.R1*$_{S1}$ oligo sample by demonstrating alkali lability (**Table A.4a**). These data prove not only that the *E. coli* MMR system can recognize and remove an rG/dT mismatch in DNA, but also that MutS and RNase HII act redundantly, via independent pathways, to remove an rG/dT mispair.

Table 4.2 Tolerance of an rG/dT mispair in *E. coli*

	LacZ.R1$_{S1}$		***LacZ.R5$_{S1}$***		***LacZ.D***	**No oligo**
Genotype[a]	Lac^+ freq.[b]	Rel. tr. freq.[c]	Lac^+ freq.	Rel. tr. freq.	Lac^+ freq.	Lac^+ freq.
WT	1.88	0.02	0.36	$4*10^{-3}$	90.7	< 0.1
	(0.82-5.97)		(0-0.48)		(38.5-177)	
rnhA	12.1	0.04	256	0.76	336	< 0.1
	(8.75-23.2)		(176-803)		(162-485)	
rnhB	21.8	0.13	0.715	$4.3*10^{-3}$	165	< 0.1
	(20.8-26.1)		(0.62-0.89)		(140-242)	
rnhA rnhB	174	0.37	22,600	48.2	470	< 0.1
	(154-196)		(19,700-27,300)		(371-2,110)	
mutS	17.4	$1.2*10^{-4}$	7.53	10^{-5}	82,300	7.76
	(14-19.6)		(5.48-14.2)		(50,000-104,950)	(4.6-9.51)
rnhA mutS	70	$4*10^{-4}$	485	$5*10^{-3}$	93,908	32.1
	(34.8-137)		(213-820)		(37,900-145,000)	(10-57.6)
rnhB mutS	33,000	0.27	5.23	10^{-5}	124,200	3.77
	(30,400-79,500)		(4.34-5.66)		(84,600-163,000)	(3.51-5.06)
rnhA rnhB mutS	145,000	0.69	22,400	0.11	212,000	24
	(33,300-320,000)		(11,600-32,600)		(100,000-267,000)	(20.2-40.2)

[a]The strains used were BW2038A,B, YSB-21A,B, YSB-22A,B, and YSB-23A,B, which were *mutS*$^+$ and BW2037A,B, YSB-19A,B, YSB-17,18, YSB-20A,B, which were *mutS*.

[b]The values are median and the range (in parenthesis) of *E. coli* Lac^+ transformant colonies reverting the missense mutation per 10^7 viable cells for each oligo. The numbers of repeats for each of the strains transformed with these oligos were 4. The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney test ($P < 0.05$) (**Table A.3b**).

[c]The relative frequencies of Lac^+ transformants (see Methods) are used for comparisons within *mutS*$^+$ or *mutS* strains.

4.4.4. Inefficient mismatch repair of an rC/dA mispair in *E. coli*

We then investigated whether MMR could recognize an rC/dA mismatch. We used an oligo of the lagging strand containing an rC ribonucleotide (*RpsL.R1*$_{S1}$) that would replace a dT in the *rpsL* wild-type gene of *E. coli* BW2037 and derivative strains (**Table A.1a**), generating the K87R mutation (*rpsL40*) that confers streptomycin resistance[99] (**Figure 4.1c**). We found that the dC/dA mismatch was efficiently repaired by the MMR system in *E. coli*, consistent with previous findings [132]. Whereas the ribonucleotide in the rC/dA mismatch in *rpsL* was removed very efficiently by RNase HII, there was only a modest, but still detectable, correction by MMR in *rnhB* cells (**Table 4.3**). These data differ from our results with an rG/dT mismatch in the *lacZ* gene, where MMR had a much stronger effect. We conclude that MMR can recognize an rC/dA mismatch, but with less efficiency than dC/dA or the complementary rG/dT mismatches.

Table 4.3 Strong effect of RNase HII and minor effect of MutS in preventing gene correction by an rC/dA mispair in *E. coli*

	RpsL.R1$_{S1}$		***RpsL.D***	**No oligo**
Genotype [a]	Str^r freq.[b]	Rel. tr. freq.[c]	Str^r freq.	Str^r freq.
WT	8	0.8	10	<0.1
	(1.46-8.60)		(5.41-17.7)	(0-0)
rnhB	1,850	316	5.86	<0.1
	(1,410-2,460)		(4.37-8.27)	(0-0.03)
mutS	11.6	0.001	10,500	0.76
	(10.5-13.1)		(8,070-12,900)	(0.03-2.70)
rnhB mutS	4,310	0.826	5,220	0.32
	(3,000-5,320)		(2,660-8,930)	(0.13-2.39)

[a]The strains used were BW2038A,B, BW2040A,B, BW2037A,B and BW2039A,B.
[b]The frequency of streptomycin resistant transformant colonies per 10^7 viable cells for wild-type, *mutS, rnhB* single mutant and *rnhB mutS* double mutant after transformation with *RpsL.R1$_{S1}$* or *RpsL.D* oligo is shown as median and 95% confidence limits (in parentheses), or alternatively range when number of repeats was <6. The numbers of repeats for each of the wt, *rnhB, mutS,* and *rnhB mutS* strains transformed with these oligos were as follows: *RpsL.R1$_{S1}$*: 4, 4, 4, 4; *RpsL.D*:4, 6, 6, 4; no oligo: 4, 4, 4, 4. The significance of all non-overlapping CI or range values was confirmed by Mann-Whitney test ($P < 0.05$) after subtraction of the background values (**Table A.3c**).
[c]Relative frequency of Str^r transformants (see Methods).

4.4.5. RNase HI, but not mismatch repair targets an RNA/DNA heteroduplex region containing an rG/dT mispair in *E. coli*

To determine if the MMR system can recognize and target an rG/dT mispair when present in an RNA/DNA heteroduplex region, we compared transformation frequencies by *LacZ.R5*$_{S1}$ with those by *LacZ.R1*$_{S1}$ and *LacZ.D* (**Figure 4.1b**). The recipient strains were derivatives of BW1947, which contains only the missense mutation in *lacZ* (nucleotide 1384 G $\rightarrow$ A), with wild-type and mutant alleles of *rnhA*, *rnhB* and *mutS* in all combinations (**Table A.1a**). Both *LacZ.R1*$_{S1}$ and *LacZ.R5*$_{S1}$ oligos produce the same rG/dT mismatch *in vivo*. In *LacZ.R5*$_{S1}$, the mismatched ribonucleotide is in a tract of 5 ribonucleotides, whereas in *LacZ.R1*$_{S1}$ it is isolated in DNA. Although the *rnhA* mutation greatly enhanced the relative transformation frequency by *LacZ.R5*$_{S1}$ in both *mutS*$^{+}$ and *mutS* cells (190 and 500-fold, respectively), *rnhB* had little effect except in combination with *rnhA* (up to 60-fold increase over the relative transformation frequency obtained in the *rnhA* single mutant strains) (**Table 4.2**). Surprisingly, *mutS* had no effect on the transformation frequency by the *LacZ.R5*$_{S1}$ oligo3, which is in stark contrast to results obtained with the *LacZ.R1*$_{S1}$ oligo (**Table 4.2**). Therefore, RNase HI, rather than RNase HII or MMR, is largely responsible for removing an RNA tract with a mismatch, whereas RNase HII and MMR are mostly responsible for processing an isolated mismatched ribonucleotide.

4.4.6. Mismatch repair of RNA/DNA mispairs in yeast cells

To test if the MMR system could recognize RNA/DNA mismatches in eukaryotic cells, we created a yeast *S. cerevisiae* strain in which we inactivated the genomic *TRP5* gene by

introducing a two-base deletion and one-base substitution to cause a frameshift and a nonsense mutation 15 bp apart (YS-301 and YS-303, **Table A.1b**). The combination of these two mutations reduced the spontaneous reversion of *trp5* to the Trp^+ phenotype to < 10^{-9}, allowing the detection of very low frequencies of gene correction despite the low efficiency of gene targeting in yeast [52, 62]. MMR was inactivated by deletion of *MSH2*, a yeast *mutS* homolog [40]. RNase H2 was inactivated by deleting the gene of the catalytic subunit, *RNH201* [20]. Unlike the *E. coli* λ recombination system [105], gene correction at the yeast *trp5* locus displayed no strand bias (**Figure A.2**). All *TRP5* oligos used here had a sequence corresponding to the *TRP5* sense strand. Wild-type and mutant cells were then transformed with an oligo containing two rNMPs to repair the deletion and another rNMP to correct the nonsense mutation in *trp5* (*TRP5.R2_R1*$_{I2\_S1}$) (**Figure 4.1d** and **Table A.2**). A DNA-only oligo (*TRP5.D*) provided a reference. The results (**Table 4.4a**) were similar to those obtained in *E. coli* with the *LacZ.R1*$_{S1}$ oligo (**Table 4.2**). Gene correction in the *rnh201* single mutant was only ~2.5-fold more than that in wild-type cells. As expected [58], the *msh2* deletion enhanced gene correction by the DNA-only oligo *TRP5.D*. With the *TRP5.R2_R1*$_{I2\_S1}$ oligo, in which both the 2-base insertion and the substitution were rNMPs, we observed a 5-fold (3.71/0.72) increase in gene correction efficiency in the *msh2* mutant and more than a 45-fold (33.8/0.72) increase in the *rnh201msh2* double mutant, making the RNA-containing oligo as efficient as the corresponding DNA-only oligo. The alkali lability of the *TRP5.R2_R1*$_{I2\_S1}$ oligo indicated that the results were not due to its contamination with DNA-only oligo (**Table A.4b**). We concluded that the yeast MMR system also recognizes RNA/DNA

mismatches and can act redundantly with RNase H2 to remove mispaired ribonucleotides embedded in DNA.

To determine whether the RNA/DNA ins/del and the rG/dT mispairs could be independently recognized by the MMR system in yeast, we constructed strains carrying either the nonsense mutation or the 2-base deletion in *trp5* and having *msh2* and/or *rnh201* mutant alleles. We transformed these strains with oligos containing either a 1-rNMP substitution (*TRP5.R1*$_{S1}$) or a 2-rNMP insertion (*TRP5.R2*$_{I2}$) (**Figure 4.1e,f** and **Table A.2**), respectively. Transformation results revealed that the rG/dT mispair was recognized by MMR in yeast (**Table A.5a**). To determine the effect of MMR on the 2-rNMP insertion we sequenced all the Trp^+ transformant clones obtained in the experiment with the *TRP5.R2*$_{I2}$ oligo in order to distinguish Trp^+ clones targeted by the oligos (*TRP5.R2*$_{I2}$ or the *TRP5.D* control) from Trp^+ revertant clones (**Tables A.5b-d**). Sequencing of Trp^+ transformants revealed that the ins/del RNA/DNA mismatch was clearly targeted both by Msh2 and RNase H2 (**Table 4.4b**).

Table 4.4 Competition between mismatch repair and RNase H type 2 functions in the removal of RNA/DNA mispairs in yeast

a

	TRP5.R2_ R1$_{l2_s1}$			***TRP5.D***		**No oligo**		**Non-specific oligo[a]**	
Genotype	Trp^+ freq.		Rel. tr. freq.[b]	Trp^+ freq.		Trp^+ freq.		Trp^+ freq.	
WT	0.72	(0.40-1.32)	0.30	2.39	(1.50-3.51)	< 0.1	(0-0)	<0.1	(0-0)
msh2	3.71	(2.07-4.31)	0.12	31.6	(24.5-37.2)	< 0.1	(0-0)	ND	ND
rnh201	1.84[c]	(1.15-2.88)	0.67	2.76	(1.83-3.29)	< 0.1	(0-0.03)	ND	ND
rnh201 msh2	33.8	(26.3-35.6)	0.76	44.6	(32.9-48.8)	< 0.1	(0-0.57)[d]	0.51	(0-1.36)

b

	TRP5.R2$_{l2}$			***TRP5.D***		**No oligo**	
Genotype	Trp^+ freq.		Rel. tr. freq.	Trp^+ freq.		Trp^+ freq.	
WT	1.65	(0.58-2.50)	0.04	39.1	(35.9-43.6)	< 0.1	(0-0)
msh2	7.64	(4.96-9.82)	0.17	46.1	(24.3-76.8)	< 0.1	(0-0)
rnh201	9.64	(7.51-15.0)	0.35	27.7	(12.3-36.9)	< 0.1	(0-0)
rnh201 msh2	26.9	(20.5-35.3)	0.38	70.0	(63.3-86.9)	< 0.1	(0-0)

(**a**) Trp^+ transformant frequencies for strains that contained both a two-base pair deletion and a nonsense mutation. (**b**) Trp^+ transformant frequencies obtained by precise correction of a two-base deletion calculated by multiplying the number of Trp^+ transformant colonies per 10^7 viable cells (shown in **Table A.5b**) by the percentage of clones with precise correction of the "CG" deletion (pattern R in **Table A.5c,d**). Each value represents the median and range (in parentheses) for 10^7 viable cells, based on 8 (a) or 4 (b) experiments. The significance of all non-overlapping confidence limits or range values was confirmed by Mann-Whitney test ($P < 0.05$) after subtraction of the no-oligo background values (**Table A.3e** and **A.3h**).

[a]A non-specific oligo, with no homology to the targeting locus, corresponding to *LEU2.R1dw* sequence (**Table A.2**), was used as additional negative control.

[b]Relative frequency of Trp^+ transformants (see Methods).

[c]The gene correction frequency value obtained in the *rnh201* mutant cells was different from that obtained in the wild-type cells at the significance level 0.1 ($P = 0.082$).

[d]Out of 7 Trp^+ revertant clones obtained after transformation with the no-oligo control in *rnh201 msh2* mutant cells none had the sequence corresponding to correction by oligos. ND, not determined.

4.5. DISCUSSION

Previously, in experiments of DSB repair in yeast using oligos containing tracts of four ribonucleotides or longer, we showed that oligos containing the shortest RNA tracts were the most efficient at chromosomal gene modification [77]. Also an oligo containing a 2-base loop within a 6-rNMP tract was 25 to 50-fold less efficient at gene correction than the corresponding DNA-only oligo in *E. coli* [67]. Contrary to expectations, it was then remarkable to find in the current work that gene correction by an oligo containing just a single rNMP is 40-fold less efficient than that obtained with an oligo containing 6 rNMPs and 2,000-fold less than that obtained with the corresponding DNA-only oligo in *E. coli* (**Table 4.1**). Here we show that the capacity of rNMPs embedded in DNA to directly transfer genetic information to the genome is not only affected by the length but also by the structure of the embedded RNA tracts, which can be targets of specific proteins and be removed before serving as templates for DNA synthesis.

Initial analysis of gene correction frequencies in *rnh* mutants in *E. coli* revealed a wide variation among different rNMP-containing oligos (**Table 4.1**). Whereas deletion of *rnhB* had no effect on the gene correction frequency by the oligo containing 5 rNMPs with the mispaired rG (*LacZ.R5*$_{S1}$) and instead it increased gene correction by the *LacZ.R6*$_{I2}$ oligo 8 fold (**Table 4.1**), RNase HII cleaved *in vitro* a corresponding substrate with the 5 rNMPs and the mispaired rG more efficiently than the substrate containing 6 rNMPs with a 2-rNMP loop (**Figure 4.2**). In the gene correction assay, RNase HI is the key factor targeting the stretch of 5 rNMPs, while RNase HII targets this substrate mainly in the absence of RNase HI (**Tables 4.1** and **4.2**). In contrast, the tract of 6 rNMPs containing

the 2-rNMP loop of the *LacZ.R6*$_{I2}$ oligo is not a good substrate for RNase HI or HII in *E. coli*. The presence of a 2-rNMP loop (generated by *LacZ.R6*$_{I2}$ or *LacZ.R2.47*$_{I2}$), but not that of a mispaired rNMP (generated by *LacZ.R1*$_{S1}$, *LacZ.R5*$_{S1}$ or *rpsL.R1*$_{S1}$) probably hampers RNase HI and RNase HII cleavage activity in *E. coli* (**Tables 4.1, 4.2** and **4.3**). Interestingly, a 2-rNMP loop is targeted by RNase H2 in yeast (**Table 4.4b**), possibly due to the processive activity of this enzyme in eukaryotes [20]. While it is known that RNase H type 1 and type 2 have distinct cleavage specificities, mostly from biochemical studies [20, 23, 126], our *in vivo* gene correction results shed light on the *in vivo* substrate specificity of these enzymes and offer an opportunity to investigate the mechanism underlying cellular tolerance of ribonucleotides misincorporated into genomic DNA.

In addition to demonstrations of ribonucleotide incorporation by DNA polymerases *in vitro* [117-121, 123], recent work from the Kunkel group provides indirect evidence that ribonucleotides are incorporated into yeast DNA by low-fidelity (*pol2-M644G*) and wild-type Pol ε alleles and can destabilize the yeast genome in Δ*rnh201* cells [111, 124], strengthening the knowledge that yeast *rnh201* mutant cells display an increased rate of genome instability [133]. Up to a 25-fold increase in the rate of spontaneous mutagenesis was observed in cells with *rnh201* null [111]. Interestingly, the frequency of base substitutions (G to A transitions) increased 6-fold in *pol2-M644G* Δ*rnh201* when the MMR gene *MSH6* was also deleted [124]. MMR could recognize RNA/DNA mismatches generated by polymerase ε and/or by other replicative polymerases [123]. Moreover, also repair polymerases, such as Pol μ, Pol β and LigD polymerase can add rNMPs in DNA [120-122], and especially error-prone polymerases might be more promiscuous at rNMP

incorporation in DNA. We suggest that RNA/DNA mismatches can arise by mispairing of an rNMP during its incorporation, or by a correct pairing during initial incorporation followed by the mispairing of a dNMP during a subsequent round of DNA synthesis. In the current work, we have found that MMR can remove the region containing a mispaired rNMP.

Among the types of RNA/DNA mismatches examined, we found that the rG/dT mispair is well recognized by the MMR system both in *E. coli* and in yeast cells. A preliminary binding experiment of yeast Msh2/Msh6 to an rG/dT mismatch [124] supports our finding that the MMR system can target this mismatch *in vivo*. The rC/dA mispair is also targeted by the MMR system in *E. coli*, although less efficiently, since gene correction by the oligo generating the rC/dA mispair is increased only a few fold in the absence of MutS in the *rnhB* mutant cells (**Table 4.3**). The sequence context surrounding the RNA/DNA mispair could certainly affect RNA/DNA mismatch recognition and/or removal. A more rigorous analysis of eukaryotic and prokaryotic MMR factor binding and ATPase functions in the context of various RNA/DNA mismatches will be an important next step to better characterize the capacity of MMR factors to process RNA/DNA mismatches. In yeast cells, gene correction by the *TRP5.R2_R1*$_{I2\_S1}$ and by the *TRP5.R2*$_{I2}$ oligos was 5-fold and 4-fold more efficient in *msh2* than in wild-type cells, respectively (**Table 4.4**). Thus, we conclude that the MMR system can target, with different specificity, RNA/DNA mismatches (in *E. coli* and yeast) or small insertions/deletions (in yeast, not tested in *E. coli*). Moreover, our work demonstrates not only that RNA/DNA mismatches are susceptible to MMR but also that MMR and RNase

H type 2 compete for RNA/DNA mismatches both in the *E. coli* and the yeast *S. cerevisiae* systems. Absence of both RNase H type 2 and MMR functions has a synergistic effect on gene correction frequency by the *TRP5.R2_R1*$_{I2\_S1}$ and the *LacZ.R1*$_{S1}$ oligos (**Table 4.4a** and **Table A.5a**).

The MMR system of *E. coli* can recognize RNA/DNA mismatches only when these consist of one or two isolated ribonucleotides embedded in DNA, but not when the RNA/DNA mismatch is within an RNA/DNA duplex region. Indeed, while the single rG/dT mispair was very efficiently removed by the MMR system, there was no detectable effect of MMR on the same rG/dT mispair in the same sequence context, when the mispair was surrounded by two rNMPs on each side (**Figure 4.1b** and **Table 4.2**). Differently from the misincorporated rNMPs scattered in DNA, which are preferred substrates for RNase H type 2 and are not targeted by RNase H type 1, mispairs in longer RNA-tracts, such as those that could be generated by error-prone primases that can misincorporate dNMPs, or those that could be present in RNA/DNA hybrid tracts of R-loops, are primarily targets of RNase H type 1, which can only be partially backed up by RNase H type 2 function.

In summary, we have demonstrated that the MMR system recognizes and targets several types of RNA/DNA mismatches present in DNA in *E. coli* and *S. cerevisiae* cells. In addition, an RNA/DNA heteroduplex region that contains a mismatch is preferentially and efficiently targeted by RNase H type 1 *in vivo*, whereas mispaired rNMPs in *E. coli* and yeast and small rNMP insertions in yeast are specific substrates for RNase H type 2.

Thus, RNase H type 2 and MMR have overlapping activity in contributing to the removal of mispaired rNMPs embedded in DNA both in a prokaryotic and a eukaryotic cell system. Our findings also open up the possibility that other DNA repair mechanisms could tackle rNMPs embedded into DNA.

4.6. ACKNOWLEDGEMENTS

The authors thank P. W. Doetsch with his group and Y. W. Kow for discussions and comments; we are grateful to G. F. Crouse and R. Pai for suggestions on the paper, C. Flood for technical assistance and all the members of the Storici laboratory for technical support and advices in the course of the study. This research was supported by the Georgia Cancer Coalition grant (R9028), the NSF MCB-1021763 grant and the Integrative Biosystems Institute grant (IBSI-4).

CONCLUSIONS

RNA molecules have a variety of important or essential biological functions in cells. In addition to mRNA, rRNA and tRNA, which play central roles in protein synthesis, a number of regulatory non-coding RNAs are involved in regulation of transcription and translation [110]. As recently demonstrated in a model organism, the yeast *S. cerevisiae*, using synthetic RNA-containing oligos, RNA can serve as a direct template for DNA synthesis during DSB repair. This was the first proof that an RNA template can directly repair a DSB and transfer genetic information to DNA without being reverse transcribed into cDNA [77].

In this study, we showed that the function of RNA as a template for DNA synthesis occurs not only in yeast but also in *E. coli* and in human HEK-293 cells. We excluded the possibility that RNA-driven repair and modification is mediated by a cDNA reverse transcript by showing that RNA-containing oligos display the same strand bias targeting as the corresponding DNA. These findings established that the capacity of RNA to directly transfer genetic information to DNA is conserved from a prokaryotic to a higher eukaryotic cell system. The direct flow of information from RNA to DNA suggests that RNA could guide genetic alterations and be a source of genetic variation.

Following the demonstration that RNA can be used as template for DNA sysnthesis at the chromosomal level in different cell types, we directed our studies to determine which RNA sequences would be more prone to transfer information to genomic DNA in cells.

For this goal, we developed a procedure to generate desired RNA/DNA hybrids at the chromosomal level by transformation using RNA-containing oligos that have homology to a specific defective reporter gene located on chromosomal DNA. If the RNA tract of the RNA-containing oligos is used as template for DNA synthesis at the chromosomal level, the defective reporter gene is corrected providing for a detectable phenotype. The frequency of reporter gene correction reflects the frequency by which a specific RNA tract can transfer information to DNA. We started this procedure in yeast *S. cereviae* and then extended it to the bacterium *E. coli*. This procedure turned out very useful especially to analyze the effect of single rNMPs or short rNMP tracts embedded in DNA, which could be frequently occurring in genomic DNA.

Recent studies have shown that rNMPs can be incorporated quite frequently into DNA by many different DNA polymerases that normally function during DNA replication and repair [25]. The misincorporated rNMPs could potentially also form base-base mispairs with the opposite dNMPs in a duplex DNA, and, if not removed, could serve as templates in the subsequent round of DNA synthesis and produce a genetic change. We postulated that the transfer of genetic information from RNA to DNA could be largely dependent on the stability of the RNA/DNA hybrids. Using our procedure to generate RNA/DNA hybrids *in vivo*, we have assessed factors that may affect the stability of RNA/DNA hybrids, such as sequence, length, structure, sequence context around the RNA tracts, and defects in RNases H and/or in DNA repair genes. We showed that isolated mispaired rNMPs in chromosomal DNA are recognized and removed not only by RNases H but also by the MMR system in both *E. coli* and yeast. Mispaired rNMPs in *E. coli* and yeast

and small rNMP insertions in yeast are specific substrates of RNase H type 2. A mismatch within an RNA/DNA heteroduplex region requires RNase H type 1 for efficient removal. Furthermore, in the absence of MMR and RNase H type 2, the RNA-driven gene modification increased dramatically. Our results suggested that RNases H could be considered as DNA repair proteins in terms of their capacity to target misincorporated rNMPs in DNA. The MMR system and RNases H play a significant role in the maintenance of genome integrity.

Overall, our results establish that the capacity of rNMPs to edit genomic DNA is conserved in three very distinct cell types, which are *E. coli*, *S. cerevisiae* and human cells, suggesting a more universal role of RNA in generating genetic variation and influencing DNA metabolism. Moreover, the finding that rNMPs are targets not only of RNases H but also of DNA repair factors, such as MMR factors both in a prokaryotic (*E. coli*) and a eukaryotic (*S. cerevisiae*) cell system, supports the fact that rNMPs are probably common non-standard nucleotides found in DNA.

In the future, in addition to the rG/dT, rC/dA and 2-nt rGrC insertion loop mispairs tested in this study, we will design RNA-containing oligos to form other types of rNMP/dNMP mispairs or RNA insertion loops to characterize more substrates for MutS/Msh2 via both *in vivo* and *in vitro* experiments. We will perform the *in vivo* oligo-driven gene correction assays in both *E. coli* and yeast systems to compare the transformation frequencies in WT and *mutS*/*msh2* mutant cells. We will generate various RNA/DNA mispair substrates by annealing RNA-containing oligos with their corresponding DNA oligos *in vitro* and test

the binding affinity of MutS/Msh2 to the RNA/DNA mismatch substrates. We will then compare results obtained with *in vivo* assays to those obtained with the *in vitro* assays for the same substrates. Other MMR factors, such as *E. coli* MutL, yeast Msh3, Msh6 and Mlh will also be tested both *in vivo* and *in vitro* to study their participation in the process of RNA/DNA mismatche repair. Besides the MMR system, other DNA repair pathways, like NER and BER will be studied to see if they are able to recognize and process misincorporated rNMPs embedded in the DNA genome.

Implications

Our results demonstrated that RNA-containing oligos or RNA-only oligos can serve as a template for DNA repair and modification at the chromosomal level in prokaryotic *E. coli* and eukaryotic yeast *S. cerevisiae* and human cells. Remarkably, when oligos were used to repair a DSB in our yeast and human cell systems, RNA-only oligos were able to repair the break and transfer the genetic information to chromosomal DNA. Differently, RNA-only oligo-directed DNA modification was not observed in *E. coli* with the Lambda Red recombination system. In the DSB repair systems the RNA-only molecules can first serve as templates for repair synthesis and then can be discarded without being incorporated into the genome. Instead, in the Lambda Red recombination system, the oligos must first be incorporated into the genome and then they serve as templates for DNA synthesis in the next round of replication. It is possible that RNA-only oligos cannot be incorporated in the *E. coli* genome because there is no ligase to ligate a 5' ribonucleotide to 3' deoxyribonucleotide. Although, we believe that RNA-only molecules could also serve as templates for DNA synthesis at the chromosomal level in *E. coli*, if they were designed to repair a DSB in the genome, as we did in our yeast and human cell system. This newly discovered capability of RNA suggests that transcript RNAs, which are abundant in the nucleus of cells, have the potential to serve as a templates to repair a DSB occurring on the chromosomal DNA during or after transcription. Transcript RNAs could be ideal templates for homology-driven DSB repair, because the sequence of transcript RNAs directly derives from that of their coding DNAs. Nascent transcript RNAs still linked to their coding DNAs in the process of transcription could serve as

templates to correct DNA lesions, such as DSBs occurring on their own parental DNAs. We speculate that highly transcribed DNA genes that experience DSBs could be repaired more efficiently than poorly or non transcribed DNAs (**Figure 6.1**). It is our future goal to set up an experimental system to detect transcript RNA-mediated DNA repair and modification in different organisms. According to our previous results, the repair frequency by synthetic RNA-only oligos was extremely low [67, 77]. Therefore, in order to observe the rare repair events by transcript RNAs, we will choose highly transcribed genes as reporters and we will also employ a DSB system to stimulate the transfer of genetic information from the RNAs to chromosomal DNAs.

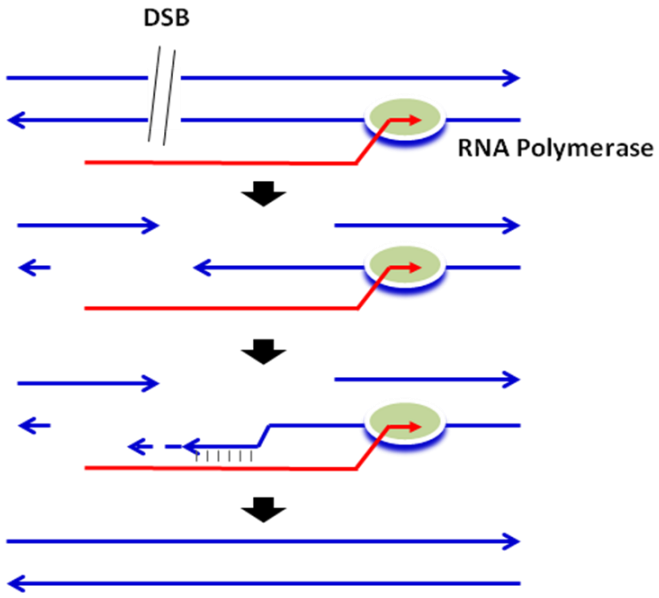


Figure 6.1 Scheme of transcript RNA-mediated DSB repair

Our results uncovered that the MMR system and RNases H can recognize and target the misincorporated rNMPs in DNA in *E. coli* and *S. cerevisiae* cells. Together with the conserved phenomenon of RNA-templated DNA repair and modification in both

prokaryotic and eukaryotic cell systems, the capacity of cells to maintain genome integrity by removing the misincorporated rNMPs is also conserved. Cells can use synthetic RNA-containing oligos for DNA synthesis and have evolved mechanisms to deal with rNMPs embedded in DNA, strengthening the hypothesis that rNMPs could be commonly present in the genome of cells. What are the biological functions of rNMPs incorporated in genomic DNA? Up to now there are only very few cases in which the presence of rNMPs in DNA is related to a specific biological function. The incorporation of one or two rNMPs was observed in the DNA duplex of the *mat1* locus of the yeast *Schizosaccharomyces pombe*, suggesting that these rNMPs can play a role in the mating-type switching in fission yeast [134, 135]. During bacterial nonhomologous end joining, rNMPs are efficiently added to DSB ends by the polymerase module of LigD enzyme, and then are resected until only a single 3'-ribonucleotide remains, which facilitate nick sealing by ligase module of the LigD enzyme [136].

It remains clearly understudied how the presence of rNMPs in genomic DNA alters DNA functions and how it affects genome integrity. There is indirect evidence that rNMPs incorporated into the yeast DNA by a low fidelity Polε (*pol2–M644G*) can cause 2-5 bp deletions in repetitive sequences at a high rate in the absence of RNase H2 [25]. It was found that if rNMPs are present at cleavage sites of Topoisomerase I, Topoisomerase I cleavage at rNMPs leads to nicks in DNA which cannot be repaired properly by Topoisomerase I and result in accumulation of deletions within short tandem repeats [137]. These frameshift mutations lead to abnormal function of protein products, therefore, are harmful to cells. Our studies revealed that isolated rNMP are targets of the

MMR system and RNase H type 2 and that a tract of rNMPs is primarily targeted by RNase H type 1. However, a 2-rNMP insertion loop cannot be recognized by RNase HII in *E. coli* and an rNMP mismatch embedded in an RNA/DNA hybrid cannot be recognized and removed by the MMR system. Our data suggest that the efficiency of rNMPs removal in DNA depends on the sequence, length, structure and sequence context of the rNMPs. The rNMPs that are not targeted by repair factors, can remain in the genome, thus could be mutagenic.

Although rNMPs embedded in DNA can cause mutations leading to abnormal biological functions, they could, potentially, also be beneficial to cells by playing a role of signaling in DNA metabolism. Differently from the methyl-directed mechanism in *E. coli*, strand discontinuity is believed to serve as a signal that directs the targeting of MMR to the discontinuous strand in eukaryotes [30]. Discontinuity exists at the termini of Okazaki fragments of the newly synthesized lagging strand during DNA replication [30]. However, it still remains unclear what could be the signal for strand discrimination by the MMR on the leading strand. Since Polε, which is the leading strand polymerases [138], appears to incorporate about 10-fold more rNMPs in DNA than Polδ [5], it was hypothesized that rNMPs incorporated by Polε after nicking by RNase H2 could serve as a signal for the recruitment of the MMR machinery to the leading strand.

Ribonucleotides in DNA can cause the DNA backbone distortion and can become a hindrance of normal B form of DNA [112]. Therefore, rNMPs in DNA could affect the protein-DNA interaction. Transcription factors can bind to DNA motifs to activate or

repress gene expression [139]. If rNMPs are incorporated into the binding sites of transcription factors, they could pose a sequence or conformational change to prevent the correct binding with transcription factors, thus altering the gene expression.

Our study demonstrated that the MMR pathway is involved in the removal of misincorporated rNMPs. There may be additional pathways for rNMP removal. The backbone distortion caused by the 2'-OH group on the ribose ring of the rNMPs could be a target of NER enzymes, especially when a tract of rNMPs is incorporated. Moreover, RNase H-dependent repair pathway for the removal of rNMPs embedded in DNA may involve FEN1 and PCNA, suggesting that rNMP repair could share common features with long-patch base excision repair [25]. Studying if other DNA repair pathways, such as NER and BER, are involved in rNMP removal can provide further insights to better understand the functions of rNMPs embedded in DNA and the significance of RNA-driven DNA modifications in cells.

Appendix A

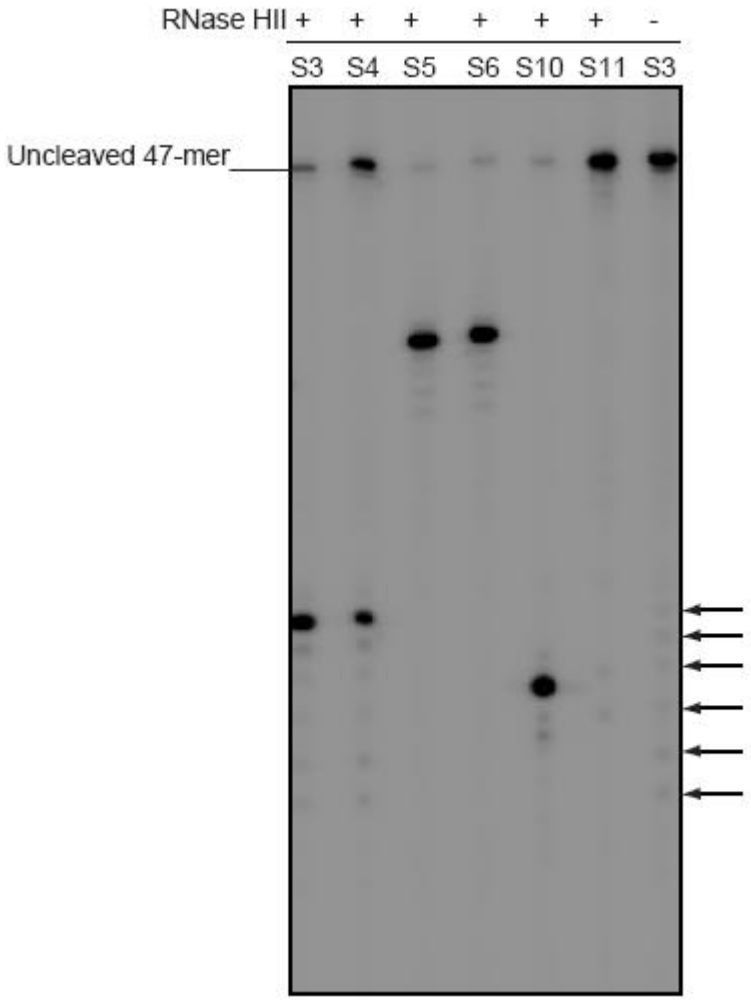
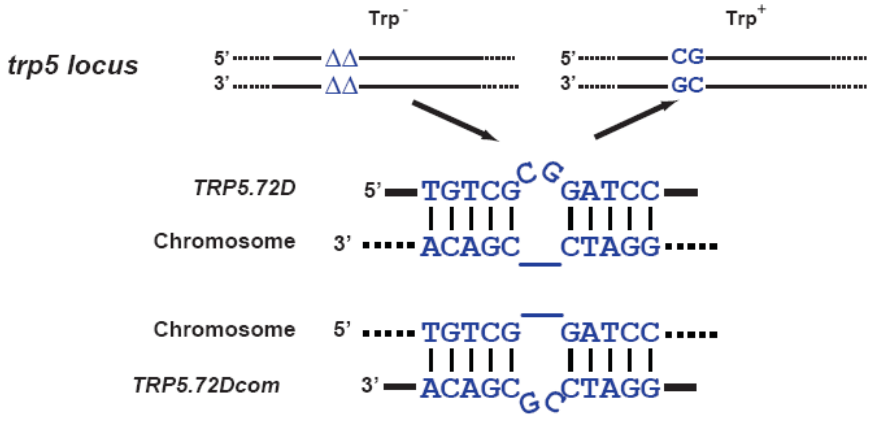


Figure A.1 Denaturing polyacrylamide gel showing fragments resulting from RNase HII cleavage and random RNA degradation in the absence of RNase HII. Structural presentation of the substrates is shown in **Figure 4.2a**. S3, the substrate that contains a stretch of 6 fully complementary ribonucleotides, was used to show fragments resulting from random RNA degradation without RNase HII. The position of the uncleaved 47-mers is shown to the left. The positions of the random cleavage of the S3 substrate in the absence of RNase HII are indicated by arrows to the right. These barely detectable degradation bands are similar to those seen for the other substrates containing more than one rNMP (this figure and **Figure 4.2b**).

a



b

Genotype	*TRP5.72D*		*TRP5.72Dcom*		No oligo	
WT	28.2	(26.6-29.6)	10.8	(8.98-11.5)	<0.1	(0-0.17)
msh2	76	(62.4-99.0)	67.9	(60.5-89.9)	5.27	(2.01-7.24)

Figure A.2 Strand-bias targeting effect of two complementary oligos, TRP5.72D and TRP5.72Dcom is mild in WT and cannot be observed in msh2 mutant strains. (a) Diagram and sequence of the chromosomal trp5 region targeted by the TRP5.72D and TRP5.72Dcom oligos. (b) Median and range of yeast Trp+ transformant colonies reverting the two-base deletion mutation per 107 viable wild-type or MMR mutant cells. The significance of all non-overlapping range values was confirmed by Mann-Whitney test ($P < 0.05$) after subtraction of the no oligo background values. Comparison of frequencies is presented in Table A.3d. The number of repeats for each of the strains transformed with the oligos was 4.

Table A.1 Bacterial and yeast strains used in this study

a

Strain[a]	Relevant genotype	Source[b]
BW1892	Δ*mutS*::*gm λcI857* Δ(*cro-bio*) *lacZ*(ΔGG1370-1; G1384→A)	[130]
BW1947	Δ*mutS*:: *gm λcI857* Δ(*cro-bio*) *lacZ*(G1384→A)	[130]
BW1947 Δ(*N-gam*)	BW1947 Δ(*N-gam*)	Transformation of BW1947 with an oligo[100]
BW1988	BW1892 Δ(*N-gam*)	Transformation of BW1892 with an oligo[100]
BW2028	BW1988 Δ*rnhA733*::*kan*	P1(JW0204-2) × BW1988
BW2029	BW1988 Δ*rnhB782*::*kan*	P1(JW0178-1) × BW1988
BW2031	BW1988 Δ*rnhB::FRT*	Excision of *kan* cassette from BW2029[140]
BW2032	BW1988 Δ*rnhB*::*FRT* Δ*rnhA733*::*kan*	P1(JW0204-2) × BW2031
BW2037A,B	Δ*mutS*:: *gm srlD3131*::Tn*10 λcI857* Δ(*cro-bio*) Δ(*N-gam*) *lacZ*(G1384→A)	P1(CAG18642) × BW1947 Δ(*N-gam*)
BW2038A,B	BW2037 *mutS*$^+$	P1(CAG18642) × BW1947 Δ(*N-gam*)
BW2039A,B	BW2037 Δ*rnhB782*::*kan*	P1(JW0178-1) × BW2037
BW2040A,B	BW2037 *mutS*$^+$ Δ*rnhB782*::*kan*	P1(JW0178-1) × BW2038
CAG18642	*srlD3131*::Tn*10*	[141, 142]
JW0178-1	Δ*rnhB782*::*kan*	[143]
JW0204-2	Δ*rnhA733*::*kan*	[143]
YSB-13,14	BW1988 *rnhA*(48Oc,Op)	Transformation of BW1988 with an oligo
YSB-15,16	BW1988 *rnhB*(70Oc,Op)	Transformation of BW1988 with an oligo
YSB-17,18	BW2037 *rnhB*(70Oc,Op)	Transformation of BW2037 with an oligo
YSB-19A,B	BW2037 *rnhA*(48Oc,Op)	Transformation of BW2037 with an oligo
YSB-20A,B	BW2037 *rnhA*(48Oc,Op) *rnhB*(70Oc,Op)	Transformation of YSB-18 with an oligo
YSB-21A,B	YSB-19 *mutS*$^+$	P1(JW2674-1) × YSB-19
YSB-22A,B	YSB-18 *mutS*$^+$	P1(JW2674-1) × YSB-18
YSB-23A,B	YSB-20 *mutS*$^+$	P1(JW2674-1) × YSB-20
JW2674-1	Δ*srlD*::(FRT-*kan*-FRT)	[143]

b

BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	[52, 144]
FRO-694	BY4742 *trp5*::GSKU	Insertion of GSKU cassette into *TRP5*
YS-301,303	BY4742 *trp5*(ΔCC1001-2; G1017→A)	Replacement of GSKU in FRO-694 with an oligo to introduce the desired mutations
YS-305,306	YS-301,303 Δ*rnh201*::*kanMX4*	Disruption of *RNH201* with *kanMX4* PCR product
YS-307,308	YS-301,303 Δ*msh2*::*kanMX4*	Disruption of *MSH2* with *kanMX4* PCR product
YS-313,314	YS-301,303 Δ*rnh201*::*hygMX4* Δ*msh2*::*kanMX4*	Disruption of *RNH201* with *hygMX4* PCR product and of *MSH2* with *kanMX4* PCR product
YS-316,318	BY4742 *trp5*(ΔCC1001-2)	Replacement of GSKU in FRO-694 with an oligo to introduce the desired mutation
YS-331,332	YS-316,318 Δ*rnh201*::*hygMX4*	Disruption of *RNH201* with *hygMX4* PCR product
YS-327,328	YS-316,318 Δ*msh2*::*kanMX4*	Disruption of *MSH2* with *kanMX4* PCR product
YS-323,324	YS-316,318 Δ*rnh201*::*hygMX4* Δ*msh2*::*kanMX4*	Disruption of *RNH201* with *hygMX4* PCR product and of *MSH2* with *kanMX4* PCR product
YS-320,322	BY4742 *trp5*(G1017→A)	Replacement of GSKU in FRO-694 with an oligo to introduce the desired mutation
YS-333,334	YS-320,322 Δ*rnh201*::*hygMX4*	Disruption of *RNH201* with *hygMX4* PCR product
YS-329,330	YS-320,322 Δ*msh2*::*kanMX4*	Disruption of *MSH2* with *kanMX4* PCR product
YS-325,326	YS-320,322 Δ*rnh201*::*hygMX4* Δ*msh2*::*kanMX4*	Disruption of *RNH201* with *hygMX4* PCR product and of *MSH2* with *kanMX4* PCR product

(**a**) *E. coli* strains. (**b**) *S. cerevisiae* strains.

[a]All *E. coli* strains are derivatives of *E. coli* K-12 F^- λ^-. Abbreviations: *gm*, gentamycin resistance gene; *kan*, kanamycin resistance gene; FRT, Flp recombinase target sequence; GSKU, cassette containing the I-SceI endonuclease gene under the inducible *GAL1* promoter, the *kanMX4* kanamycin resistance gene, the counterselectable *KlURA3* marker and the I-SceI cutting site.

[b]Transductions with bacteriophage P1 are described as follows: P1(donor) × recipient. Selections were for resistance to kanamycin, tetracycline (Tn*10*), or gentamycin.

Table A.2 Oligos used in this study

Name	Size	Structure from 5'	Sequence
LacZ.R6$_{I2}$	65	[D 22 R 2] ins::R 2 [R 2 D 10] sD1 [D 26]	5'-GTAATCACCCGAGTGTGATCATrCrU**rGrGr**UrCGCTGGGGAAT**G**AGTCAGGCCACGGCGCTAATCACGAC
LacZ.R1$_{S1}$	65	[D 24] ins::D 2 [D 12] sR1 [D 26]	5'-GTAATCACCCGAGTGTGATCATCT**GG**TCGCTGGGGAAT**rG**AGTCAGGCCACGGCGCTAATCACGAC
LacZ.R5$_{S1}$	65	[D 24] ins::D 2 [D 10 R 2] sR1 [R 2 D 24]	5'-GTAATCACCCGAGTGTGATCATCT**GG**TCGCTGGGGArArU**rG**rArGTCAGGCCACGGCGCTAATCACGAC
LacZ.D	65	[D 24] ins::D 2 [D 12] sD1 [D 26]	5'-GTAATCACCCGAGTGTGATCATCT**GG**TCGCTGGGGAAT**G**AGTCAGGCCACGGCGCTAATCACGAC
RpsL.R1$_{S1}$	70	[D 34] sR 1 [D 35]	5'- ACGTACGGTGTGGTAACGAACACCCGGGAGGTCT**rC**TAACACGACCGCCACGGATCAGGATCACGGAGTGC
RpsL.D	70	[D 34] sD 1 [D 35]	5'- ACGTACGGTGTGGTAACGAACACCCGGGAGGTCT**C**TAACACGACCGCCACGGATCAGGATCACGGAGTGC
TRP5.R2_R1$_{I2\_S1}$	65	[D 24] ins::R 2 [D 14] sR1 [D 24]	5'- AAAAGGGTTTTGATGAAGCTGTCG**rCrG**GATCCCACATTCTG**rG**GAAGACTTCAAATCCTTGTATTCT
TRP5.D	65	[D 24] ins::D 2 [D 14] sD1 [D 24]	5'- AAAAGGGTTTTGATGAAGCTGTCG**CG**GATCCCACATTCTG**GG**AAGACTTCAAATCCTTGTATTCT
TRP5.R1$_{S1}$	65	[D 40] sR1 [D 24]	5'- AAAAGGGTTTTGATGAAGCTGTCGCCGATCCCACATTCTG**rG**GAAGACTTCAAATCCTTGTATTCT
TRP5.R2$_{I2}$	65	[D 24] ins::R 2 [D 39]	5'- AAAAGGGTTTTGATGAAGCTGTCG**rCrG**GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
TRP5.Dcc	65	[D 40] sD1 [D 24]	5'- AAAAGGGTTTTGATGAAGCTGTCGCCGATCCCACATTCTG**GG**AAGACTTCAAATCCTTGTATTCT
TRP5.72D	72	[D 35] ins::D 2 [D 35]	5'-AAGAGAGTTGGAAAAGGGTTTTGATGAAGCTGTCG**CG**GATCCCACATTCTGGGAAGACTTCAAATCCTTGTA
TRP5.72Dcom	72	[D 35] ins::D 2 [D 35]	5'-TACAAGGATTTGAAGTCTTCCCAGAATGTGGGATC**CG**CGACAGCTTCATCAAAACCCTTTTCCAACTCTCTT
LEU2.R1dw	60	D 44 R1 D15	5'-TTAGGTGCTGTGGGTGGTCCTAAATGGGGATCCGGTAGTGTTAGrGCCTGAACAAGGTTTA
LacZ.R1.47	47		5'-CCCGAGTGTGATCATCTGGTCGCTGGGGAATrGAGTCAGGCCACGGCG
LacZ.R6.47	47		5'-CCCGAGTGTGATCATrCrUrGrGrUrCGCTGGGGAATGAGTCAGGCCACGGCG
LacZ.R2.47$_{I2}$	47	[D 17] ins::R 2 [D 28]	5'-CCCGAGTGTGATCATCT**rGrG**TCGCTGGGGAATGAGTCAGGCCACGGCG
LacZ.R5.47	47		5'-CCCGAGTGTGATCATCTGGTCGCTGGGGArArUrGrArGTCAGGCCACGGCG
LacZ.D.47	47	[D 17] ins::D 2 [D 28]	5'-CCCGAGTGTGATCATCT**GG**TCGCTGGGGAATGAGTCAGGCCACGGCG
LacZ.comD.wt.47	47		5'-CGCCGTGGCCTGACTCATTCCCCAGCGACCAGATGATCACACTCGGG
LacZ.comD.m1.47	47		5'-CGCCGTGGCCTGACTTATTCCCCAGCGACCAGATGATCACACTCGGG
LacZ.comD.del2.45	45		5'-CGCCGTGGCCTGACTCATTCCCCAGCGAAGATGATCACACTCGGG

In the name of the RNA-containing oligos, substitutions are indicated by a subscript capital "S" and insertions by a subscript capital "I". The letters "S" and "I" are followed by a subscript number indicating the number of bases that are substituted or inserted, respectively. The structures of the oligos used in this study are described from the 5'end, with DNA sequences (D) shown in blue and RNA sequences (R) in red. Bases with homology to the chromosomal DNA are shown in brackets and underlined. Insertions are indicated as "ins::". Base substitutions are indicated as "s". In the oligo sequences, the base changes introduced by the oligos are in bold and shown in blue, if these consist of DNA bases and in red, with lower case "r" on the left side, if these consist of RNA bases. The RNA tracts of the non-specific *LEU2.R1dw* oligo and of the *lacZ* oligos used only in the RNase HII cleavage assay are shown in regular red type. The restriction sites that are introduced by the oligos are underlined (thin underline for *Van*91I, thick line for *Bam*HI). The DNA oligos were desalted (synthesized by Invitrogen, Carlsbad, CA, or Alpha DNA, Montreal, Quebec, Canada); the RNA–containing oligos were desalted and deprotected (by Thermo Scientific Dharmacon, Lafayette, CO). The *LacZ.R1*$_{S1}$ and all the *lacZ* 45- and 47-mers were also PAGE purified (by Thermo Scientific Dharmacon or by Invitrogen).

Table A.3 Statistical comparisons (P-values) between gene correction frequencies obtained for different oligos in different genetic backgrounds

a

I)

Oligo	*mutS vs. mutS rnhA*	*mutS vs. mutS rnhB*	*mutS vs. mutS rnhA rnhB*	*mutS rnhA vs. mutS rnhA rnhB*	*mutS rnhB vs. mutS rnhA rnhB*
LacZ.R6$_{l2}$	0.0960	0.0012	0.0021	0.0021	0.0034
LacZ.R2.47$_{l2}$	ND	0.6857	ND	ND	ND
LacZ.R1$_{S1}$	0.0032	0.0021	0.0034	0.0034	0.6166
LacZ.R5$_{S1}$	0.0050	0.7000	0.0022	0.0050	0.0022

II)

Oligo	*mutS*	*mutS rnhA*	*mutS rnhB*	*mutS rnhA rnhB*
LacZ.R6$_{l2}$ vs. LacZ.R2.47$_{l2}$	0.0061	ND	0.7619	ND
LacZ.R6$_{l2}$ vs. LacZ.R1$_{S1}$	0.0021	0.0021	0.0034	0.0034
LacZ.R6$_{l2}$ vs. LacZ.R5$_{S1}$	0.0167	0.0214	0.0238	0.0034
LacZ.R1$_{S1}$ vs. LacZ.R5$_{S1}$	0.0222	0.0218	0.0222	0.0022
LacZ.R1$_{S1}$ vs. LacZ.R2.47$_{l2}$	0.0106	ND	0.0106	ND
LacZ.R5$_{S1}$ vs. LacZ.R2.47$_{l2}$	0.0095	ND	0.0095	ND

Mann-Whitney test was applied to determine whether a difference exists between various pairs of gene correction frequencies. (**a**) Comparison of frequencies presented in **Table 4.1**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between relative frequencies obtained for each oligo in different backgrounds, and II) between relative frequencies obtained for different oligos in the same backgrounds. ND, not determined.

b
I)

Oligo	WT *vs.* *rnhA*	WT *vs. rnhB*	WT *vs.* *rnhA rnhB*	*rnhA vs.* *rnhA rnhB*	*rnhB vs.* *rnhA rnhB*
LacZ.R1$_{S1}$	0.0286	0.0286	0.0286	0.0286	0.0286
LacZ.R5$_{S1}$	0.0294	0.0294	0.0294	0.0286	0.0286
LacZ.D	0.0571	0.1143	0.0286	0.3429	0.0286

Oligo	*mutS vs.* *rnhA mutS*	*mutS vs.* *rnhB mutS*	*mutS vs.* *rnhA rnhB mutS*	*rnhA mutS vs.* *rnhA rnhB mutS*	*rnhB mutS vs.* *rnhA rnhB mutS*
LacZ.R1$_{S1}$	0.0286	0.0286	0.0286	0.0286	0.2
LacZ.R5$_{S1}$	0.0286	0.6857	0.0286	0.0286	0.0286
LacZ.D	1.0000	0.1143	0.1143	0.1143	0.2

Oligo	WT *vs.* *mutS*	*rnhA vs.* *rnhA mutS*	*rnhB vs.* *rnhB mutS*	*rnhA rnhB vs. rnhA* *rnhB mutS*
LacZ.R1$_{S1}$	0.0286	0.0286	0.0286	0.0286
LacZ.R5$_{S1}$	1.0000	1.0000	0.3429	1.0000
LacZ.D	0.0286	0.0286	0.0286	0.0286

II)

Oligo	WT	*rnhA*	*rnhB*	*rnhA rnhB*
No oligo *vs. LacZ.R1*$_{S1}$	0.0286	0.0286	0.0286	0.0286
No oligo vs. *LacZ.R5*$_{S1}$	1	0.0286	0.1143	0.0286

(**b**) Comparison of frequencies presented in **Table 4.2**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between frequencies obtained for each oligo in different backgrounds, and II) between frequencies obtained for different oligos in the same backgrounds.

c

I)

Oligo	WT *vs. rnhB*	WT *vs. mutS*	WT *vs. rnhB mutS*	*rnhB vs. rnhB mutS*	*mutS vs. rnhB mutS*
RpsL.R1$_{S1}$	0.029	0.029	0.029	0.029	0.029
RpsL.D	0. 171	0.010	0.029	0.010	0. 038

II)

Oligo	WT	*rnhB*	*mutS*	*rnhB mutS*
No oligo *vs. RpsL.R1$_{S1}$*	0.021	0.027	0.029	0.029
No oligo *vs. RpsL.D*	0.021	0.013	0.010	0.029
RspL.R1$_{S1}$ vs. RpsL.D	0.486	0.010	0.010	0.486

(**c**) Comparison of frequencies presented in **Table 4.3**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between frequencies obtained for each oligo in different backgrounds, and II) between frequencies obtained for different oligos in the same backgrounds.

d

I)

Oligo	WT *vs. msh2*
TRP5.72D	0.029
TRP5.72Dcomp	0.029
No oligo	0.027

II)

Oligo	WT	*msh2*
No oligo *vs. TRP5.72D*	0.027	0.029
No oligo *vs. TRP5.72Dcomp*	0.027	0.029
TRP5.72D vs. TRP5.72Dcomp	0.029	0.886

(**d**) Comparison of frequencies presented in **Figure A.2b**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between frequencies obtained for each oligo in different backgrounds, and II) between frequencies obtained for different oligos in the same backgrounds.

e

I)

Oligo	**WT *vs.* *msh2***	**WT *vs.* *rnh201***	**WT *vs.* *rnh201 msh2***	***msh2* vs. *rnh201 msh2***	***rnh201*vs. *rnh201 msh2***
TRP5.R2_R1$_{l2_S2}$	0.008	0.082	0.001	0.001	0.001
TRP5.D	0.001	0. 875	0.001	0.083	0.001
Non-specific oligo	ND	ND	0.197	ND	ND

II)

Oligo	**WT**	***msh2***	***rnh201***	***rnh201 msh2***
No oligo *vs.TRP5.R2_R1$_{l2_S2}$*	0.004	0.001	<0.001	0.001
No oligo *vs. TRP5.D*	0.001	<0.001	<0.001	0.001
TRP5.R2_R1$_{l2_S2}$ vs. TRP5.D	0.023	0.002	0. 267	0.052
No oligo *vs.* Non-specific oligo	ND	ND	ND	0.814

(**e**) Comparison of frequencies presented in **Table 4.4a**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between frequencies obtained for each oligo in different backgrounds, and II) between frequencies obtained for different oligos in the same backgrounds. ND, not determined.

f

I)

Oligo	WT *vs. msh2*	WT *vs. rnh201*	WT *vs.* *rnh201 msh2*	*msh2 vs.* *rnh201 msh2*	*rnh201 vs.* *rnh201 msh2*
TRP5.R1$_{S1}$	0.057	0. 486	0.029	0. 114	0.029
TRP5.Dcc	0.029	0. 486	0.029	0. 343	0.029

II)

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*
No oligo *vs.TRP5.R1*$_{S1}$	0.028	0.029	0.028	0.029
No oligo *vs*. *TRP5.Dcc*	0.029	0.029	0.029	0.029
TRP5.R1$_{S1}$ *vs*. *TRP5.Dcc*	0.307	0.029	0.029	1.000

(**f**) Comparison of frequencies presented **Table A.5a**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between frequencies obtained for each oligo in different backgrounds, and II) between frequencies obtained for different oligos in the same backgrounds.

g

I)

Oligo	WT *vs. msh2*	WT *vs. rnh201*	WT *vs. rnh201 msh2*	*msh2 vs. rnh201 msh2*	*rnh201 vs. rnh201 msh2*
TRP5.R2$_{l2}$	0. 114	0.029	0.029	0.029	0.029
TRP5.D	0.029	0.057	0.029	0.2	0.029

II)

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*
No oligo *vs.TRP5.R2$_{l2}$*	0.041	0.029	0.029	0.029
No oligo *vs. TRP5.D*	0.029	0.029	0.029	0.029
TRP5.R2$_{l2}$ vs. TRP5.D	0.029	0.029	0.057	0.029

(**g**) Comparison of frequencies presented in **Table A.5b**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between frequencies obtained for each oligo in different backgrounds, and II) between frequencies obtained for different oligos in the same backgrounds.

h

I)

Oligo	WT *vs. msh2*	WT *vs. rnh201*	WT *vs. rnh201 msh2*	*msh2 vs. rnh201 msh2*	*rnh201 vs. rnh201 msh2*
TRP5.R2$_{l2}$	0. 029	0.029	0.029	0.029	0.029
TRP5.D	1	0.057	0.029	0.342	0.029

II)

Oligo	WT	*msh2*	*rnh201*	*rnh201 msh2*
TRP5.R2$_{l2}$ vs. TRP5.D	0.029	0.029	0.057	0.029

(**h**) Comparison of frequencies presented in **Table 4.4b**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. I) Comparisons were between frequencies obtained for each oligo in different backgrounds, and II) between frequencies obtained for different oligos in the same backgrounds.

i

Oligo	BW-2028 *vs.* YSB-13, 14	BW-2029 *vs.* YSB-15, 16
LacZ.R6$_{I2}$	1	0.548
LacZ.R1$_{S1}$	0.109	0.647

(**i**) Comparison of frequencies presented in **Table A.6a**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. Comparisons were between frequencies obtained for each oligo in different backgrounds.

j

Oligo	BW2039 *vs*. YSB-18
LacZ.R1$_{S1}$	0.710
LacZ.D	0.111
No oligo	0.114

(**j**) Comparison of frequencies presented in **Table A.6b**. Two groups in a pair were considered to be significantly different when P-values were less than 0.05. Comparisons were between frequencies obtained for each oligo in different backgrounds.

Table A.4 Alkali lability of RNA-containing oligos used in *E. coli* or yeast transformation assays

a

Oligo	WT		*mutS*		*rnhB*		*rnhB mutS*	
LacZ.R1$_{S1}$ w/o NaOH	1.98	(1.98-2.96)	16.5	(15.0-18.1)	7.84	(3.72-7.84)	25,100	(21,600-29,300)
LacZ.R6$_{I2}$ w/o NaOH	ND		583	(394-772)	ND		ND	
LacZ.R2.47$_{I2}$ w/o NaOH	ND		821	(204-1,020)	ND		ND	
LacZ.D w/o NaOH	156	(131-186)	49,400	(7,840-50,200)	60.3	(51.3-81.4)	78,400	(53,500-100,900)
LacZ.D.47 w/o NaOH	ND		1,990	(1,990-3,670)	ND		ND	
LacZ.R1$_{S1}$ w/ NaOH	0.25	(0-0.25)	10.5	(8.63-11.8)	<0.1	(0-0.1)	91.2	(75.4-120)
LacZ.R6$_{I2}$ w/ NaOH	ND		0.62	(0.31-0.63)	ND		ND	
LacZ.R2.47$_{I2}$ w/ NaOH	ND		17.2	(8.01-17.4)	ND		ND	
LacZ.D w/ NaOH	98.8	(91.4-126)	36,900	(32,900-52,500)	29.2	(19.1-36.2)	47,800	(25,200-91,700)
LacZ.D.47 w/ NaOH	ND		2,090	(1,940-2,140)	ND		ND	
No oligo	<0.1		9.41		<0.1		18.3	(13.3-23.4)
No oligo (BW1988)	ND		<0.1		ND		ND	

b

Oligo	WT		*rnh201 msh2*	
TRP5.R2_R1$_{I2\_S1}$ w/o NaOH	0.19	(0-0.373)	10.7	(8.46-13.2)
TRP5.D w/o NaOH	4.34	(3.05-5.22)	38.4	(37.2-42.3)
TRP5.R2_R1$_{I2\_S1}$ w/ NaOH	<0.1	(0-0)	0.255	(0-0.510)
TRP5.D w/ NaOH	1.7	(1.11-3.44)	28.5	(28.0-33.8)
No oligo	<0.1	(0-0)	<0.1	(0-0)

c

Oligo	*rnh201 msh2*	
TRP5.R1$_{S1}$ w/o NaOH	81.9	(77.1-83.5)
TRP5.Dcc w/o NaOH	110	(108-128)
TRP5.R1$_{S1}$ w/ NaOH	5.62	(4.82-10.4)
TRP5.Dcc w/ NaOH	97.2	(78.7-101)
No oligo	17. 7	(16.1-29.7)

d

Oligo	*rnh201 msh2*	
TRP5.R2$_{I2}$ w/o NaOH	37.8	(30.9-43.0)
TRP5.D w/o NaOH	84.2	(59.3-99.7)
TRP5.R2$_{I2}$ w/ NaOH	4.3	(1.72-6.01)
TRP5.D w/ NaOH	96.2	(87.6-125)
No oligo	2.58	(2.58-5.15)

(**a**) Median and range (in parentheses) of *E. coli* Lac^+ transformant colonies, per 10^7 viable wild-type, MMR and RNase H mutant cells, following treatment or no treatment with NaOH, obtained using *LacZ.R1*$_{S1}$, *LacZ.R6*$_{I2}$, *LacZ.R2.47*$_{I2}$, *LacZ.D* or *LacZ.D.47*. The *LacZ.5R*$_{S1}$ and *RpsL.R1*$_{S1}$ oligos were not tested by alkaline hydrolysis, since it was clear from data in **Table 4.2** and **Table 4.3**, respectively, that the RNA-containing oligo solutions were indeed containing RNA and were not contaminated by the corresponding DNA-only oligo solutions. (**b**) Median and range (in parentheses) of yeast Trp^+ transformant colonies, per 10^7 viable wild-type, RNase H and MMR mutant cells, following treatment or no treatment with NaOH, obtained to correct the two-base deletion and nonsense mutations, (**c**) only the nonsense mutation, or (**d**) only the two-base deletion mutation. ND, not determined.

Table A.5 Reversion frequency of a nonsense mutation or a two-base deletion in the yeast *trp5* gene following transformation by rNMP-containing oligos in MMR and RNase H mutant cells

a

Oligo	WT		*msh2*		*rnh201*		*rnh201 msh2*	
TRP5.R1$_{S1}$	9.01	(8.23-11.9)	31.4[a]	(20.6-43.7)	11.2	(9.23-12.6)	84.7	(35.1-130)
TRP5.Dcc	7.79	(5.46-11.9)	56.4	(51.4-69.1)	7.52	(5.93-7.84)	78.6	(54.6-90.9)
No oligo	0.68	(0-0.99)	10.7	(8.6-19.3)	0.68	(0.37-1.65)	14.8	(8.2-20.7)

b

Oligo	WT		*msh2*		*rnh201*		*rnh201 msh2*	
TRP5.R2$_{I2}$	3.54 **46.7%**	(1.24-5.35)	12.6[b] **60.6%**	(8.18-16.2)	9.64 **100%**	(7.51-15.0)	35.5 **75.7%**	(27.1-46.6)
TRP5.D	39.1 **100%**	(35.9-43.6)	52.3 **88.2%**	(27.6-87.0)	27.7 **100%**	(12.3-36.9)	84.6 **82.8%**	(76.5-105)
No oligo	0.38 **0.0%**	(0-1.24)	5.27 **0.0%**	(2.01-7.24)	0.28 **0.0%**	(0-1.70)	6.53 **0.0%**	(3.73-9.80)

c

Pattern	Sequence
Chr	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
R	AAAAGGGTTTTGATGAAGCTGTCG**CG**GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
R1	AAAAGGGTTTTGATGAAGCTGTCG**CGA**ATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
R2	AAAAGGGTTTTGATGAAGCTGTCG**CG**GA**C**CCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
A	AAAAGGGTTT-GATGAAGCTGTCG--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
B	AAAAGGGTTTTGATGAAGCTGTC---GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
C	AAAAGGGTTTTGATGAAGCTGTCG--GATCC-ACATTCTGGGAAGACTTCAAATCCTTGTATTCT
D	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCACATTCTGG-AAGACTTCAAATCCTTGTATTCT
E	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCA-ATTCTGGGAAGACTTCAAATCCTTGTATTCT
F	AAAAGGGTTTTGATGAAGCTGTC--GGATCCCACATTCTGGGAAGACTTCAAATCCTTG-ATTCT
G	AAAAGGGTTTTGATGAA----TC**T**--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
H	AAAAGGGTTTTGATGAAGCTGTCG--G-TCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
I	AAAAGGGTTTT**TG**GATGAAGCTGTCG--GATCCCACATTCTGGGAAGACTTCAAATCCTTGTATTCT
J	AAAAGGGTTTTGATGAAGCTGTCG--GATCCCACATA-TGGGAAGACTTCAAATCCTTGTATTCT

d

Oligo	WT		*msh2*		*rnh201*		*rnh201 msh2*	
No oligo	R	0/3	R	0/14	R	0/4	R	0/18
	A	1/3	A	10/14	A	2/4	A	11/18
	C	1/3	D	3/14	B	1/4	C	2/18
	D	1/3	E	1/14	C	1/4	E	5/18
TRP5.R2$_{I2}$	R	7/15	R	20/33	R	65/65	R	115/152
	B	7/15	A	9/33			R1	1/152
	H	1/15	B	1/33			R2	1/152
			C	1/33			A	28/152
			E	1/33			C	1/152
			I	1/33			D	2/152
							F	1/152
							G	2/152
							J	1/152
TRP5.D	R	28/28	R	15/17	R	30/30	R	24/29
			A	2/17			A	5/29

(**a**) Median and range (in parentheses) of Trp^+ transformant colonies per 10^7 viable wild-type, MMR and RNase H mutant cells, obtained to revert the nonsense mutation, or (**b**) the two-base deletion in *trp5*. The number of repeats for each of the strains transformed with these oligos was 4. Results obtained for oligos *TRP5.R1$_{S1}$* and *TRP5.R2$_{I2}$* were not due to contamination with DNA oligos (**Table A.4c**,**d**). The significance of all non-overlapping CI or range values was confirmed by Mann-Whitney test ($P < 0.05$) after subtraction of the no oligo background values. Comparison of frequencies is presented in

Tables A.3f and A.3g. In (b), the percentage of clones with precise correction of the "CG" deletion (pattern R in (c)) is shown in bold under the median. The median and these percentages were used to calculate the frequency of Trp^+ transformant colonies with precise correction of the two-base deletion mutation presented in (**Table 4.4b**).

(**c**) Sequence patterns of the *TRP5* region in Trp^+ colonies transformed with no oligo, *TRP5.R2$_{I2}$*, or *TRP5.D* oligos. Differences from the chromosomal sequence are in bold. Chr, chromosomal sequence of *trp5* with the 2-base deletion mutation; R, sequence of the *TRP5* region targeted with oligo *TRP5.R2$_{I2}$* or *TRP5.D*, containing the "CG" insertion; R1, R2, the *TRP5* region corrected by oligos presenting additional mutations; A-J, patterns of the *TRP5* region from spontaneous Trp^+ revertants. (**d**) The frequency of each sequence pattern obtained in WT, *msh2*, *rnh201* and *rnh201 msh2* colonies following sequence analysis of 408 Trp^+ clones from the experiment shown in (b) above. These data were used to generate results presented in **Table 4.4b**.

[a]The gene correction frequency value obtained in the *msh2* mutant cells was different from that obtained in the wild-type cells at the significant level 0.1 (P = 0.057) after subtraction of the no-oligo background value.

[b]The gene correction frequency value obtained in the *msh2* mutant cells was not significantly different from that obtained in the wild-type cells (P = 0.114) after subtraction of the no-oligo background value. However, after sequencing all the Trp^+ clones of this experiment, the frequency value of precise gene correction obtained in the *msh2* mutant cells was significantly different from that obtained in the wild-type cells (P = 0.029, see **Table 4.4b** and **Table A.3h**).

Table A.6 Deletion and nonsense mutations of *rnhA* and *rnhB* genes similarly affect gene correction by oligos in the BW1988 and the BW2037 backgrounds

a

	LacZ.R6$_{I2}$		***LacZ.R1$_{S1}$***	
***rnhA* (BW2028)**	0.04	(0.02-0.05)	$9*10^{-4}$	($7*10^{-4}$-$1.4*10^{-3}$)
***rnhA* (YSB-13, 14)**	0.04	(0.02-0.05)	$7*10^{-4}$	($5.5*10^{-4}$-$9.1*10^{-4}$)
***rnhB* (BW2029)**	0.16	(0.08-0.19)	3.43	(1.74-3.87)
***rnhB* (YSB-15, 16)**	0.12	(0.09-0.17)	3.07	(2.67-3.97)

b

	LacZ. R1$_{S1}$		***LacZ.D***	**No oligo**
Genotype	Lac^+ freq.	Rel. tr. freq.[a]	Lac^+ freq.	Lac^+ freq.
***rnhB mutS* (BW2039)**	17,500	0.2	86,700	3.98
	(6,250-30,000)		(45,000-103,000)	(2.23-25.4)
***rnhB mutS* (YSB-18)**	33,000	0.27	124,200	3.77
	(30,400-79,500)		(84,600-163,000)	(3.51-5.06)

(**a**) The strains used are BW2028, YSB-13, 14, BW2029, and YSB-15, 16, all of which are *mutS* mutants. All the strains were transformed with the *LacZ.R6$_{I2}$*, *LacZ.R1$_{S1}$*, or *LacZ.D* oligo. The values are relative transformation frequencies (see Methods). The numbers of repeats for each of strains transformed with these oligos were as follows: *LacZ.R6$_{I2}$:* 7,3,7,3; *LacZ.R1$_{S1}$*:6,3,7,3; *LacZ.D*: 6,3,6,3. The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney test ($P < 0.05$) (**Table A.3i**). (**b**) The strains used were BW2039 and YSB-18. The strains were transformed with *LacZ.R1$_{S1}$* and *LacZ.D* oligos. The values are median and the range (in parenthesis) of Lac^+ transformant colonies reverting the missense mutation per 10^7 cells. The numbers of repeats for each of the strains transformed with these oligos were 4. The significance of all non-overlapping confidence limit values was confirmed by the Mann-Whitney test ($P < 0.05$) (**Table A.3j**).

[a]Relative frequency of Lac^+ transformants (see Methods). In the absence of a transforming oligo, the number of Lac^+ transformants per 10^7 viable cells were <0.1.

REFERENCES

1. Hoeijmakers, J.H., *Genome maintenance mechanisms for preventing cancer.* Nature, 2001. **411**(6835): p. 366-74.
2. Peterson, C.L. and J. Cote, *Cellular machineries for chromosomal DNA repair.* Genes Dev, 2004. **18**(6): p. 602-16.
3. Aguilera, A. and B. Gomez-Gonzalez, *Genome instability: a mechanistic view of its causes and consequences.* Nat Rev Genet, 2008. **9**(3): p. 204-17.
4. Joyce, C.M., *Choosing the right sugar: how polymerases select a nucleotide substrate.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1619-22.
5. Nick McElhinny, S.A., et al., *Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases.* Proc Natl Acad Sci U S A, 2010. **107**(11): p. 4949-54.
6. Nick McElhinny, S.A. and D.A. Ramsden, *Polymerase mu is a DNA-directed DNA/RNA polymerase.* Mol Cell Biol, 2003. **23**(7): p. 2309-15.
7. Kim, N. and S. Jinks-Robertson, *dUTP incorporation into genomic DNA is linked to transcription in yeast.* Nature, 2009. **459**(7250): p. 1150-3.
8. Rowen, L. and A. Kornberg, *A ribo-deoxyribonucleotide primer synthesized by primase.* J Biol Chem, 1978. **253**(3): p. 770-4.
9. Kuchta, R.D. and G. Stengel, *Mechanism and evolution of DNA primases.* Biochim Biophys Acta, 2009.
10. Randerath, K., et al., *Formation of ribonucleotides in DNA modified by oxidative damage in vitro and in vivo. Characterization by 32P-postlabeling.* Mutat Res, 1992. **275**(3-6): p. 355-66.
11. Egli, M., N. Usman, and A. Rich, *Conformational influence of the ribose 2'-hydroxyl group: crystal structures of DNA-RNA chimeric duplexes.* Biochemistry, 1993. **32**(13): p. 3221-37.
12. Stein, H. and P. Hausen, *Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase.* Science, 1969. **166**(903): p. 393-5.
13. Kanaya, S. and M. Ikehara, *Functions and structures of ribonuclease H enzymes.* Subcell Biochem, 1995. **24**: p. 377-422.
14. Shuck, S.C., E.A. Short, and J.J. Turchi, *Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology.* Cell Res, 2008. **18**(1): p. 64-72.
15. Parikh, S.S., et al., *Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5083-8.
16. Ohtani, N., et al., *Identification of the genes encoding Mn2+-dependent RNase HII and Mg2+-dependent RNase HIII from Bacillus subtilis: classification of RNases H into three families.* Biochemistry, 1999. **38**(2): p. 605-18.
17. Qiu, J., et al., *Saccharomyces cerevisiae RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease.* Mol Cell Biol, 1999. **19**(12): p. 8361-71.

18. Turchi, J.J., et al., *Enzymatic completion of mammalian lagging-strand DNA replication.* Proc Natl Acad Sci U S A, 1994. **91**(21): p. 9803-7.
19. Ohtani, N., et al., *Molecular diversities of RNases H.* J Biosci Bioeng, 1999. **88**(1): p. 12-9.
20. Cerritelli, S.M. and R.J. Crouch, *Ribonuclease H: the enzymes in eukaryotes.* FEBS J, 2009. **276**(6): p. 1494-505.
21. Cerritelli, S. and R. Crouch, *Ribonuclease H: the enzymes in eukaryotes.* FEBS J, 2009. **276**(6): p. 1494-505.
22. Ohtani, N., M. Tomita, and M. Itaya, *Junction ribonuclease: a ribonuclease HII orthologue from Thermus thermophilus HB8 prefers the RNA-DNA junction to the RNA/DNA heteroduplex.* Biochem J, 2008. **412**(3): p. 517-26.
23. Eder, P.S., R.Y. Walder, and J.A. Walder, *Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA.* Biochimie, 1993. **75**(1-2): p. 123-6.
24. Hou, J., et al., *RNase HII from Chlamydia pneumoniae discriminates mismatches incorporation into DNA-rN1-DNA/DNA duplexes.* Biochem Biophys Res Commun, 2007. **356**(4): p. 988-92.
25. Nick McElhinny, S.A., et al., *Genome instability due to ribonucleotide incorporation into DNA.* Nat Chem Biol, 2010. **6**(10): p. 774-81.
26. Frank, P., C. Braunshofer-Reiter, and U. Wintersberger, *Yeast RNase H(35) is the counterpart of the mammalian RNase HI, and is evolutionarily related to prokaryotic RNase HII.* FEBS Lett, 1998. **421**(1): p. 23-6.
27. Champoux, J. and S. Schultz, *Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription.* FEBS J, 2009. **276**(6): p. 1506-16.
28. Crow, Y., et al., *Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection.* Nat Genet, 2006. **38**(8): p. 910-6.
29. Stephenson, J.B., *Aicardi-Goutieres syndrome (AGS).* Eur J Paediatr Neurol, 2008. **12**(5): p. 355-8.
30. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.
31. Fukui, K., *DNA mismatch repair in eukaryotes and bacteria.* J Nucleic Acids, 2010. **2010**.
32. Jiricny, J., *The multifaceted mismatch-repair system.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 335-46.
33. Kunkel, T.A., *Nucleotide repeats. Slippery DNA and diseases.* Nature, 1993. **365**(6443): p. 207-8.
34. Costantino, N. and D.L. Court, *Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15748-53.
35. Su, S.S., et al., *Mispair specificity of methyl-directed DNA mismatch correction in vitro.* J Biol Chem, 1988. **263**(14): p. 6829-35.
36. Modrich, P., *Methyl-directed DNA mismatch correction.* J Biol Chem, 1989. **264**(12): p. 6597-600.

37. Langle-Rouault, F., G. Maenhaut-Michel, and M. Radman, *GATC sequences, DNA nicks and the MutH function in Escherichia coli mismatch repair.* EMBO J, 1987. **6**(4): p. 1121-7.
38. Mechanic, L.E., B.A. Frankel, and S.W. Matson, *Escherichia coli MutL loads DNA helicase II onto DNA.* J Biol Chem, 2000. **275**(49): p. 38337-46.
39. Burdett, V., et al., *In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6765-70.
40. Iyer, R.R., et al., *DNA mismatch repair: functions and mechanisms.* Chem Rev, 2006. **106**(2): p. 302-23.
41. Harfe, B.D., B.K. Minesinger, and S. Jinks-Robertson, *Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast.* Curr Biol, 2000. **10**(3): p. 145-8.
42. Schofield, M.J. and P. Hsieh, *DNA mismatch repair: molecular mechanisms and biological function.* Annu Rev Microbiol, 2003. **57**: p. 579-608.
43. Li, G.M., *DNA mismatch repair and cancer.* Front Biosci, 2003. **8**: p. d997-1017.
44. McCulloch, S.D., L. Gu, and G.M. Li, *Nick-dependent and -independent processing of large DNA loops in human cells.* J Biol Chem, 2003. **278**(50): p. 50803-9.
45. Modrich, P., *Mechanisms in eukaryotic mismatch repair.* J Biol Chem, 2006. **281**(41): p. 30305-9.
46. Dimitri, A., et al., *Transcription elongation past O6-methylguanine by human RNA polymerase II and bacteriophage T7 RNA polymerase.* Nucleic Acids Res, 2008. **36**(20): p. 6459-71.
47. Hazra, T.K., et al., *Characterization of a novel 8-oxoguanine-DNA glycosylase activity in Escherichia coli and identification of the enzyme as endonuclease VIII.* J Biol Chem, 2000. **275**(36): p. 27762-7.
48. Moerschell, R.P., S. Tsunasawa, and F. Sherman, *Transformation of yeast with synthetic oligonucleotides.* Proc Natl Acad Sci U S A, 1988. **85**(2): p. 524-8.
49. Ellis, H.M., et al., *High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6742-6.
50. Campbell, C.R., et al., *Homologous recombination involving small single-stranded oligonucleotides in human cells.* New Biol, 1989. **1**(2): p. 223-7.
51. Pierce, E.A., et al., *Oligonucleotide-directed single-base DNA alterations in mouse embryonic stem cells.* Gene Ther, 2003. **10**(1): p. 24-33.
52. Storici, F., L.K. Lewis, and M.A. Resnick, *In vivo site-directed mutagenesis using oligonucleotides.* Nat Biotechnol, 2001. **19**(8): p. 773-6.
53. Murphy, K.C., *Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli.* J Bacteriol, 1998. **180**(8): p. 2063-71.
54. Yu, D., et al., *An efficient recombination system for chromosome engineering in Escherichia coli.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5978-83.
55. Zhang, Y., et al., *DNA cloning by homologous recombination in Escherichia coli.* Nat Biotechnol, 2000. **18**(12): p. 1314-7.
56. Karakousis, G., et al., *The beta protein of phage lambda binds preferentially to an intermediate in DNA renaturation.* J Mol Biol, 1998. **276**(4): p. 721-31.

57. Muniyappa, K. and C.M. Radding, *The homologous recombination system of phage lambda. Pairing activities of beta protein.* J Biol Chem, 1986. **261**(16): p. 7472-8.
58. Kow, Y.W., et al., *Oligonucleotide transformation of yeast reveals mismatch repair complexes to be differentially active on DNA replication strands.* Proc Natl Acad Sci U S A, 2007. **104**(27): p. 11352-7.
59. Storici, F. and M.A. Resnick, *Delitto perfetto targeted mutagenesis in yeast with oligonucleotides.* Genet Eng (N Y), 2003. **25**: p. 189-207.
60. Vasquez, K.M., et al., *Manipulating the mammalian genome by homologous recombination.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8403-10.
61. Mandecki, W., *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis.* Proc Natl Acad Sci U S A, 1986. **83**(19): p. 7177-81.
62. Storici, F., et al., *Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 14994-9.
63. Banga, S.S. and J.B. Boyd, *Oligonucleotide-directed site-specific mutagenesis in Drosophila melanogaster.* Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1735-9.
64. Smih, F., et al., *Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells.* Nucleic Acids Res, 1995. **23**(24): p. 5012-9.
65. Storici, F. and M.A. Resnick, *The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast.* Methods Enzymol, 2006. **409**: p. 329-345.
66. Radecke, F., et al., *Targeted chromosomal gene modification in human cells by single-stranded oligodeoxynucleotides in the presence of a DNA double-strand break.* Mol Ther, 2006. **14**(6): p. 798-808.
67. Shen, Y., et al., *RNA-driven genetic changes in bacteria and in human cells.* Mutat Res, 2011.
68. Paques, F. and J.E. Haber, *Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 349-404.
69. Baltimore, D., *Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome.* Cell, 1985. **40**(3): p. 481-2.
70. Autexier, C. and N. Lue, *The structure and function of telomerase reverse transcriptase.* Annu Rev Biochem, 2006. **75**: p. 493-517.
71. Moore, J. and J. Haber, *Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks.* Nature, 1996. **383**(6601): p. 644-6.
72. Teng, S., B. Kim, and A. Gabriel, *Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks.* Nature, 1996. **383**(6601): p. 641-4.
73. Pomerantz, R. and M. O'Donnell, *The replisome uses mRNA as a primer after colliding with RNA polymerase.* Nature, 2008. **456**(7223): p. 762-6.
74. Nowacki, M., et al., *RNA-mediated epigenetic programming of a genome-rearrangement pathway.* Nature, 2008. **451**(7175): p. 153-8.
75. Kasahara, M., et al., *RecA protein-dependent R-loop formation in vitro.* Genes Dev, 2000. **14**(3): p. 360-5.

76. Huertas, P. and A. Aguilera, *Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination.* Mol Cell, 2003. **12**(3): p. 711-21.
77. Storici, F., et al., *RNA-templated DNA repair.* Nature, 2007. **447**(7142): p. 338-41.
78. Boeke, J.D., C.A. Styles, and G.R. Fink, *Saccharomyces cerevisiae SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements.* Mol Cell Biol, 1986. **6**(11): p. 3575-81.
79. Smogorzewska, A. and T. de Lange, *Regulation of telomerase by telomeric proteins.* Annu Rev Biochem, 2004. **73**: p. 177-208.
80. Storici, F., et al., *Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing.* Mol Cell Biol, 2006. **26**(20): p. 7645-57.
81. Beelman, C.A. and R. Parker, *Degradation of mRNA in eukaryotes.* Cell, 1995. **81**(2): p. 179-83.
82. LaCava, J., et al., *RNA degradation by the exosome is promoted by a nuclear polyadenylation complex.* Cell, 2005. **121**(5): p. 713-24.
83. Baltimore, D., *Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome.* Cell, 1985. **40**(3): p. 481-2.
84. Autexier, C. and N.F. Lue, *The structure and function of telomerase reverse transcriptase.* Annu. Rev. Biochem., 2006. **75**: p. 493-517.
85. Teng, S.C., B. Kim, and A. Gabriel, *Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks.* Nature, 1996. **383**(6601): p. 641-4.
86. Moore, J.K. and J.E. Haber, *Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks.* Nature, 1996. **383**(6601): p. 644-6.
87. Derr, L.K. and J.N. Strathern, *A role for reverse transcripts in gene conversion.* Nature, 1993. **361**(6408): p. 170-3.
88. Nevo-Caspi, Y. and M. Kupiec, *cDNA-mediated Ty recombination can take place in the absence of plus-strand cDNA synthesis, but not in the absence of the integrase protein.* Current Genetics, 1997. **32**(1): p. 32-40.
89. Lesage, P. and A.L. Todeschini, *Happy together: the life and times of Ty retrotransposons and their hosts.* Cytogenetic & Genome Research, 2005. **110**(1-4): p. 70-90.
90. Morrish, T.A., et al., *DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.* Nature Genetics, 2002. **31**(2): p. 159-65.
91. Storici, F., *RNA-mediated DNA modifications and RNA-templated DNA repair.* Curr Opin Mol Ther, 2008. **10**(3): p. 224-30.
92. Kasahara, M., et al., *RecA protein-dependent R-loop formation in vitro.* Genes & Development, 2000. **14**(3): p. 360-5.
93. Huertas, P. and A. Aguilera, *Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination.* Molecular Cell, 2003. **12**(3): p. 711-21.
94. Ohsato, T., et al., *R-Loop in the replication origin of human mitochondrial DNA is resolved by RecG, a Holliday junction-specific helicase.* Biochem Biophys Res Commun, 1999. **255**(1): p. 1-5.

95. Yasukawa, T., et al., *Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand.* Embo J, 2006. **25**(22): p. 5358-71.
96. Chalker, D.L., *Dynamic nuclear reorganization during genome remodeling of Tetrahymena.* Biochim Biophys Acta, 2008. **1783**(11): p. 2130-6.
97. Garnier, O., et al., *RNA-mediated programming of developmental genome rearrangements in Paramecium tetraurelia.* Mol Cell Biol, 2004. **24**(17): p. 7370-9.
98. Shen, Y. and F. Storici, *Generation of RNA/DNA hybrids in genomic DNA by transformation using RNA-containing oligonucleotides.* J Vis Exp, 2010(45).
99. Weiss, B., *Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation.* DNA Repair (Amst), 2008. **7**(2): p. 205-12.
100. Ellis, H.M., et al., *High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(12): p. 6742-6.
101. Porteus, M.H. and D. Baltimore, *Chimeric nucleases stimulate gene targeting in human cells.* Science, 2003. **300**(5620): p. 763.
102. Hirsch, M.L., et al., *AAV Recombineering with Single Strand Oligonucleotides.* Plos One, 2009. **4**(11): p. -.
103. Grieger, J.C., V.W. Choi, and R.J. Samulski, *Production and characterization of adeno-associated viral vectors.* Nat Protoc, 2006. **1**(3): p. 1412-28.
104. Lim, D. and W.K. Maas, *Reverse-Transcriptase Dependent Synthesis of a Covalently Linked, Branched DNA Rna Compound in Escherichia-Coli B.* Cell, 1989. **56**(5): p. 891-904.
105. Li, X.T., et al., *Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli.* Nucleic Acids Res, 2003. **31**(22): p. 6674-87.
106. Urnov, F.D., et al., *Highly efficient endogenous human gene correction using designed zinc-finger nucleases.* Nature, 2005. **435**(7042): p. 646-51.
107. Radecke, F., et al., *Targeted chromosomal gene modification in human cells by single-stranded oligodeoxynucleotides in the presence of a DNA double-strand break.* Molecular Therapy, 2006. **14**(6): p. 798-808.
108. Mathias, S.L., et al., *Reverse-Transcriptase Encoded by a Human Transposable Element.* Science, 1991. **254**(5039): p. 1808-1810.
109. Houseley, J., J. LaCava, and D. Tollervey, *RNA-quality control by the exosome.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 529-39.
110. Chu, C.Y. and T.M. Rana, *Small RNAs: Regulators and guardians of the genome.* Journal of Cellular Physiology, 2007. **213**(2): p. 412-419.
111. Nick McElhinny, S.A., et al., *Genome instability due to ribonucleotide incorporation into DNA.* Nat Chem Biol. , 2010.
112. Rydberg, B. and J. Game, *Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16654-9.

113. Roberts, R.W. and D.M. Crothers, *Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition.* Science, 1992. **258**(5087): p. 1463-6.
114. Mehler, M.F., *Epigenetics and the nervous system.* Ann Neurol, 2008. **64**(6): p. 602-17.
115. Mattick, J.S. and M.F. Mehler, *RNA editing, DNA recoding and the evolution of human cognition.* Trends Neurosci, 2008. **31**(5): p. 227-33.
116. Storici, F. and M.A. Resnick, *The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast.* Methods Enzymol, 2006. **409**: p. 329-45.
117. Patel, P.H. and L.A. Loeb, *Multiple amino acid substitutions allow DNA polymerases to synthesize RNA.* J Biol Chem, 2000. **275**(51): p. 40266-72.
118. Astatke, M., et al., *A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3402-7.
119. Bonnin, A., et al., *A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase.* J Mol Biol, 1999. **290**(1): p. 241-51.
120. Nick McElhinny, S.A. and D.A. Ramsden, *Polymerase mu is a DNA-directed DNA/RNA polymerase.* Molecular & Cellular Biology, 2003. **23**(7): p. 2309-15.
121. Cavanaugh, N.A., W.A. Beard, and S.H. Wilson, *DNA polymerase beta ribonucleotide discrimination: insertion, misinsertion, extension, and coding.* J Biol Chem, 2010. **285**(32): p. 24457-65.
122. Gong, C., et al., *Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C.* Nat Struct Mol Biol, 2005. **12**(4): p. 304-12.
123. Nick McElhinny, S.A., et al., *Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases.* Proc Natl Acad Sci U S A. **107**(11): p. 4949-54.
124. Clark, A.B., et al., *Mismatch repair-independent tandem repeat sequence instability resulting from ribonucleotide incorporation by DNA polymerase varepsilon.* DNA Repair (Amst), 2011. **10**(5): p. 476-82.
125. Kim, N., et al., *Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I.* Science, 2011. **332**(6037): p. 1561-4.
126. Tadokoro, T. and S. Kanaya, *Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes.* FEBS J, 2009. **276**(6): p. 1482-93.
127. Itaya, M. and R.J. Crouch, *Correlation of activity with phenotypes of Escherichia coli partial function mutants of rnh, the gene encoding RNase H.* Molecular and General Genetics, 1991. **227**(3): p. 433-7.
128. Karu, A.E., et al., *The gamma protein specified by bacteriophage lambda. Structure and inhibitory activity for the recBC enzyme of Escherichia coli.* Journal of Biological Chemistry, 1975. **250**(18): p. 7377-87.
129. Sternberg, N.L. and R. Maurer, *Bacteriophage-mediated generalized transduction in Escherichia coli and Salmonella typhimurium.* Methods Enzymol. , 1991. **204**: p. 18-24.
130. Weiss, B., *Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation.* DNA Repair, 2008. **7**: p. 205-212.

131. Harfe, B.D. and S. Jinks-Robertson, *DNA mismatch repair and genetic instability.* Annu Rev Genet, 2000. **34**: p. 359-399.
132. Dohet, C., R. Wagner, and M. Radman, *Repair of defined single base-pair mismatches in Escherichia coli.* Proc Natl Acad Sci U S A, 1985. **82**(2): p. 503-505.
133. Arudchandran, A., et al., *The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair.* Genes Cells, 2000. **5**(10): p. 789-802.
134. Vengrova, S. and J.Z. Dalgaard, *RNase-sensitive DNA modification(s) initiates S. pombe mating-type switching.* Genes Dev, 2004. **18**(7): p. 794-804.
135. Vengrova, S. and J.Z. Dalgaard, *The wild-type Schizosaccharomyces pombe mat1 imprint consists of two ribonucleotides.* EMBO Rep, 2006. **7**(1): p. 59-65.
136. Zhu, H. and S. Shuman, *Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide.* J Biol Chem, 2008. **283**(13): p. 8331-9.
137. Kim, N., et al., *Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I.* Science, 2011. **332**(6037): p. 1561-4.
138. Pursell, Z.F., et al., *Yeast DNA polymerase epsilon participates in leading-strand DNA replication.* Science, 2007. **317**(5834): p. 127-30.
139. Latchman, D.S., *Transcription factors: an overview.* Int J Biochem Cell Biol, 1997. **29**(12): p. 1305-12.
140. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6640-5.
141. Nichols, B.K., O. Shapiq, and V. Meiners, *Sequence analysis of Tn10 insertion sites in a collection of Escherichia coli strains used for genetic mapping and strain construction* J. Bacteriol. , 1998. **180**: p. 6408-6411.
142. Singer, M., et al., *A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of Escherichia coli* Microbiol. Rev. , 1989. **53**: p. 1-24.
143. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Mol. Sys. Biol., 2006. **2**: p. 2006.0008.
144. Brachmann, C.B., et al., *Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.* Yeast, 1998. **14**(2): p. 115-32.